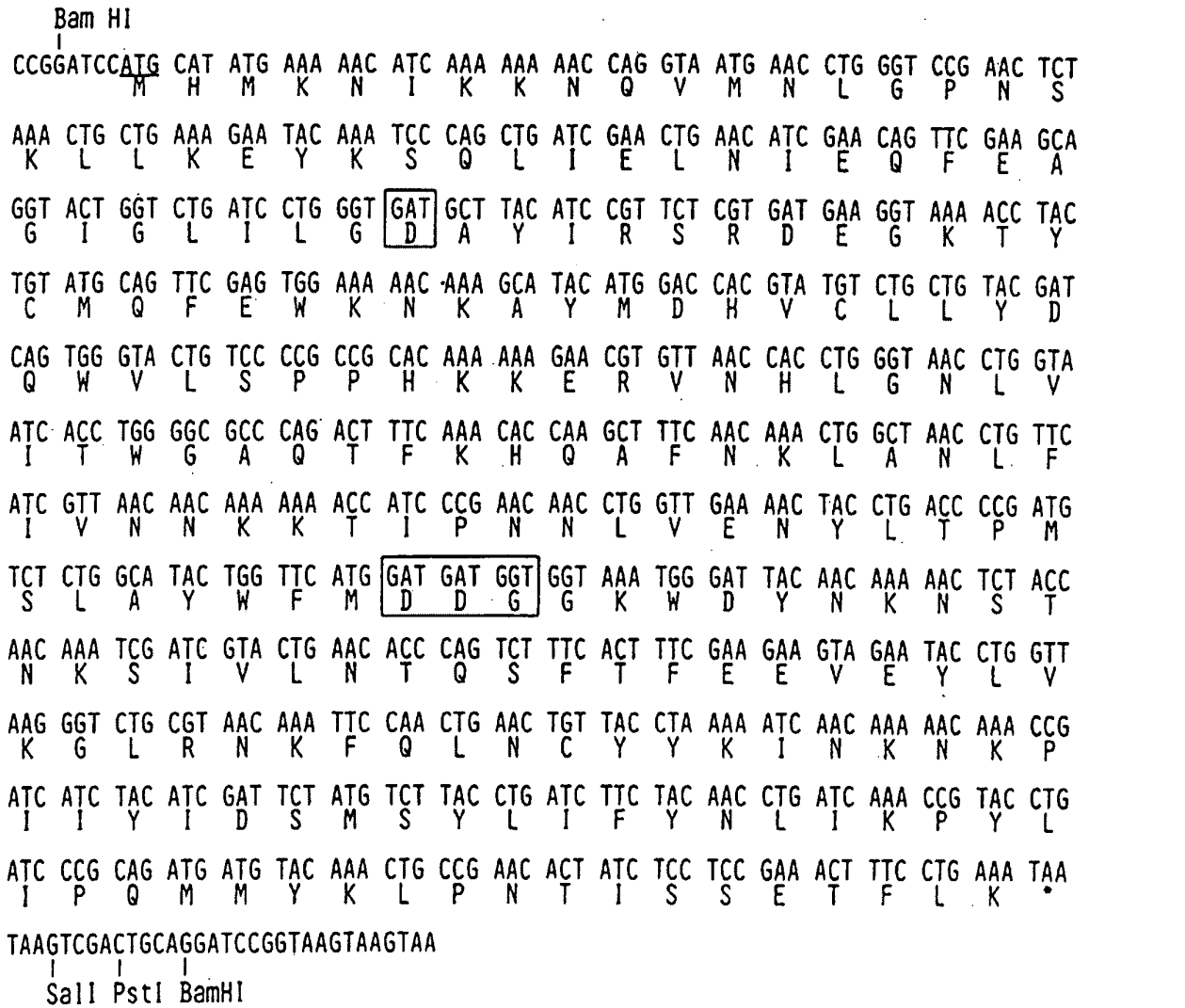
```
AAAAATAAAATCAT ATG AAA AAT ATT AAA AAA AAT CAA GTA ATC AAT CTC GGT CCT AAT TCT
               M   K   N   I   K   K   N   Q   V   M   N   L   G   P   N   S

AAA TTA TTA AAA GAA TAT AAA TCA CAA TTA ATT GAA TTA AAT ATT GAA CAA TTT GAA GCA
K   L   L   K   E   Y   K   S   Q   L   I   E   L   N   I   E   Q   F   E   A

GGT ATT GGT TTA ATT TTA GGA GAT GCT TAT ATT CGT AGT CGT GAT GAA GGT AAA ACT TAT
G   I   G   L   I   L   G   D   A   Y   I   R   S   R   D   E   G   K   T   Y

TGT ATG CAA TTT CAC TCC AAA AAT AAG GCA TAC ATG GAT CAT GTA TGT TTA TTA TAT GAT
C   M   Q   F   E   W   K   N   K   A   Y   M   D   H   V   C   L   L   Y   D

CAA TGG GTA TTA TCA CCT CCT CAT AAA AAA GAA AGA GTT AAT CAT TTA GGT AAT TTA GTA
Q   W   V   L   S   P   P   H   K   K   E   R   V   N   H   L   G   N   L   V

ATT ACC TGG GGA GCT CAA ACT TTT AAA CAT CAA GCT TTT AAT AAA TTA GCT AAC TTA TTT
I   T   W   G   A   Q   T   F   K   H   Q   A   F   N   K   L   A   N   L   F

ATT GTA AAT AAT AAA AAA CTT ATT CCT AAT AAT TTA GTT GAA AAT TAT TTA ACA CCT ATG
I   Y   N   N   K   K   L   I   P   N   N   L   V   E   N   Y   L   T   P   M

AGT CTG GCA TAT TGG TTT ATG GAT GAT GGA GGT AAA TGG GAT TAT AAT AAA AAT TCT CTT
S   L   A   Y   W   F   M   D   D   G   G   K   W   D   Y   N   K   N   S   L

AAT AAA AGT ATT GTA TTA AAT ACA CAA AGT TTT ACT TTT GAA GAA GTA GAA TAT TTA CTT
N   K   S   I   V   L   N   T   Q   S   F   T   F   E   E   V   C   Y   L   V

AAA GGT TTA AGA AAT AAA TTT CAA TTA AAT TGT TAT GTT AAA ATT AAT AAA AAT AAA CCA
K   G   L   R   N   K   F   Q   L   N   C   Y   V   K   I   N   K   N   K   P

ATT ATT TAT ATT GAT TCT AGT AGT TAT CTG ATT TTT TAT AAT TTA ATT AAA CCT TAT TTA
I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N   I   T   K   P   Y   L

ATT CCT CAA ATG ATG TAT AAA CTG CCT AAT ACT ATT TCA TCC GAA ACT TTT TTA AAA TAA
I   P   Q   M   M   Y   K   L   P   N   T   I   S   S   E   T   F   L   K   *
```

FIG. 1



1 and 2: THESE AMINO ACIDS ARE ABSOLUTELY NECESSARY TO PRODUCE CATALYTIC ACTIVITY. OTHER SUBSTITUTIONS ARE POSSIBLE, SUCH AS DELETIONS OF THE 10 FIRST AMINO ACIDS.

FIG. 2

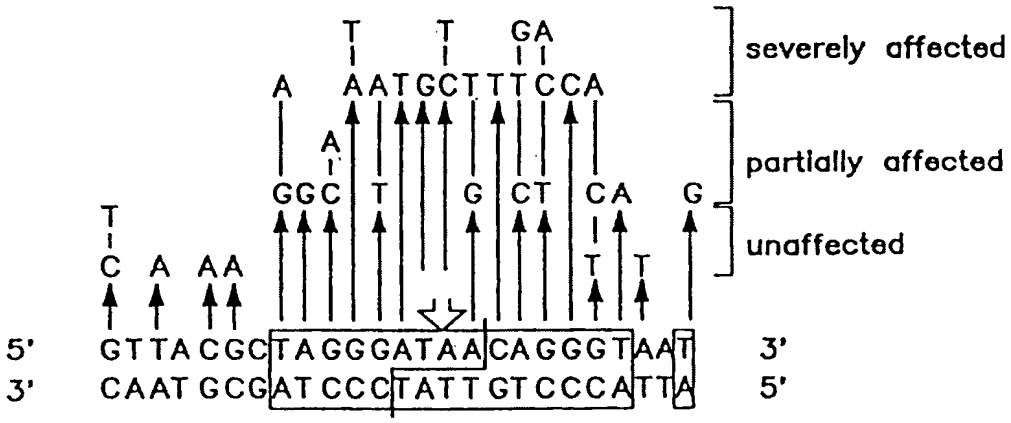


FIG. 3

```
1667 GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT 1746

1747 TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTG ATG CTC GTC AGG GGG GCG GAG  1818
   1                                                     M   L   V   R   G   A   E    7

1819 CCT ATG GAA AAA CGC CAG CAA CGC GGC CTT TTT ACG GTT CCT GGC CTT TTG CTG GCC TTT  1878
   8 P   M   E   K   R   Q   Q   R   G   L   F   T   V   P   G   L   L   L   A   F   27

1879 TGC TCA CAT GTT CTT TCC TGC GTT ATC CCC TGA TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC 1947
  28 C   S   H   V   L   S   C   V   I   P   *                                       38

1948 TGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAC 2027

2028 CGCCTCTCCCCGCGCGTTGGCCGATTCATTA ATG CAG CTG GCA CGA CAG GTT TCC CGA CTG GAA AGC  2094
   1                                 M   Q   L   A   R   Q   V   S   R   L   E   S   12

2095 GGG CAG TGA GCGCAACGCAATTA ATG TGA GTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTT ATG    2164
  13 G   Q   *                   M   *                                       M      1

2165 CTT CCG GCT CGT ATG TTG TGT GGA ATT GTG AGC GGA TAA CAATTTCACACAGGAAACAGCT ATG   2228
   2 L   P   A   R   M   L   C   G   I   V   S   G   *                           M    1

2229 ACC ATG ATT ACG AAT TCT CAT GTT TGA CAGCTTATCATCGATAAGCTTTA ATG CGG TAG TTTATCAC 2295
   2 T   M   I   T   N   S   H   V   *                       M   R   *               3

2296 AGTTAAATTGCTAACGCAGTCAGGCACCGTGT ATG AAA TCT AAC AAT GCG CTC ATC GTC ATC CTC GGC 2363
   1                                 M   K   S   N   N   A   L   I   V   I   L   G   12

2364 ACC GTC ACC CTG GAT GCT GTA GGC ATA GGC TTG GTT ATG CCG GTA CTG CCG GGC CTC TTG  2423
  13 T   V   T   L   D   A   V   G   I   G   L   V   M   P   V   L   P   G   L   L   32

2424 CGG GAT ATC CGC CTG ATG CGT GAA CGT GAC GGA CGT AAC CAC CGC GAC ATG TGT GTG CTG  2483
  33 R   D   I   R   L   M   R   E   R   D   G   R   N   H   R   D   M   C   V   L   52

2484 TTC CGC TGG GCA TGC CAG GAC AAC TTC TGG TCC GGT AAC GTG CTG AGC CCG GCC AAG CTT  2543
  53 F   R   W   A   C   Q   D   N   F   W   S   G   N   V   L   S   P   A   K   L   72
```

FIG. 4A

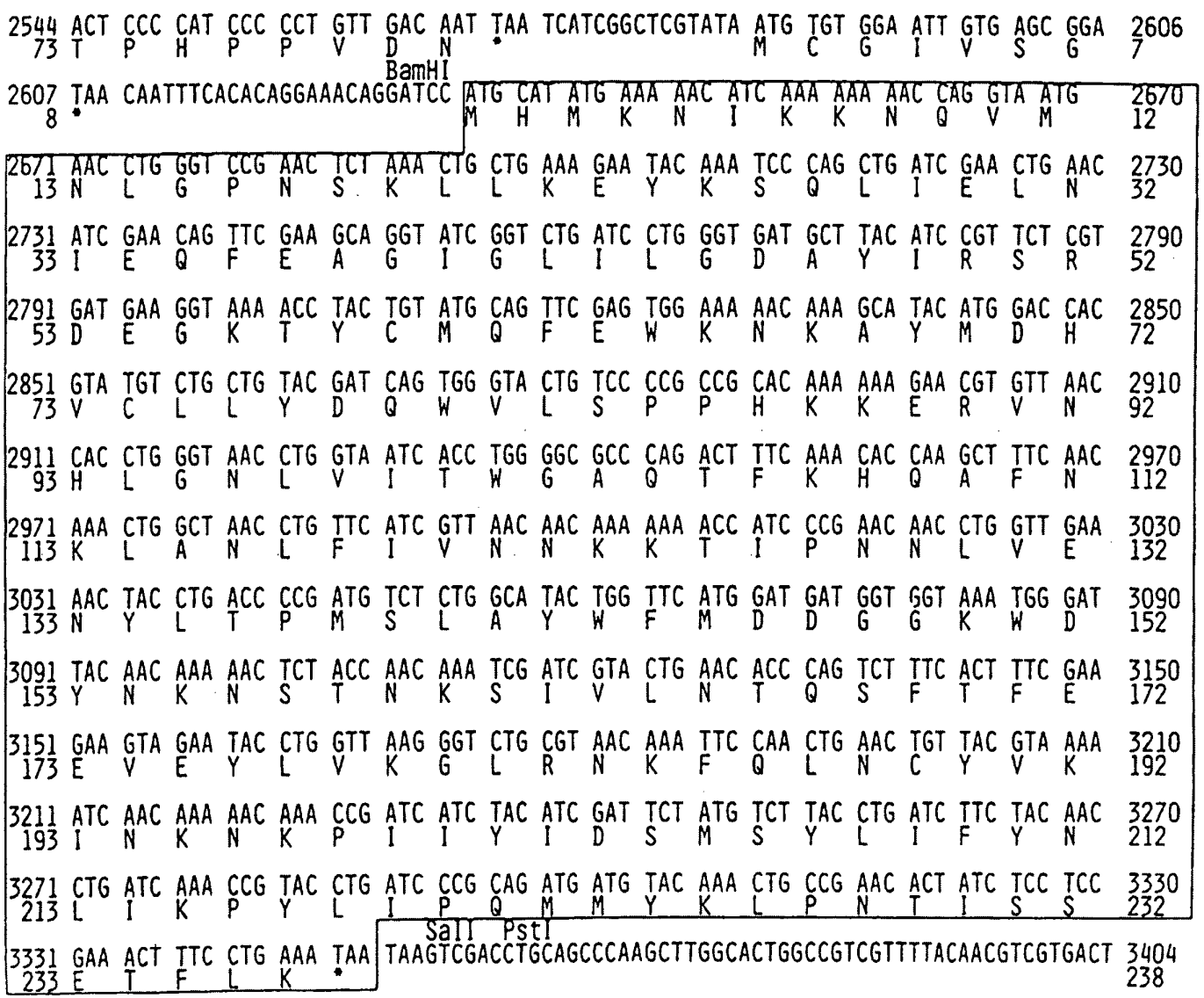


FIG. 4B

		-2	-1	1				5					10						
		M	H	M	K	N	I	K	K	N	Q	V	M	N	L	G	P	N	S
			20										30						
K	L	L	K	E	Y	K	S	Q	L	I	E	L	N	I	E	Q	F	E	A
			40										50						
G	I	G	L	I	L	G	D	A	Y	I	R	S	R	D	E	G	K	T	Y
			60										70						
C	M	Q	F	E	W	K	N	K	A	Y	M	D	H	V	C	L	L	Y	C
			80										90						
Q	W	Y	L	S	P	P	H	K	K	E	R	Y	N	H	L	G	N	L	Y
			100										110						
I	T	W	G	A	Q	T	F	K	H	Q	A	F	N	K	L	A	N	L	F
			120										130						
I	V	N	N	K	K	T	I	P	N	N	L	V	E	N	Y	L	T	P	M
			140										150						
G	L	A	Y	W	P	M	D	D	G	G	K	W	D	Y	N	K	N	S	I
			160										170						
N	K	S	I	V	L	N	T	Q	S	F	T	F	E	E	V	E	Y	L	V
			180										190						
K	G	L	R	N	K	F	Q	L	N	C	Y	V	K	I	N	K	N	K	P
			200										210						
I	I	Y	I	D	S	M	S	Y	L	I	F	Y	N	L	I	K	P	Y	L
			220										230						
I	P	Q	M	M	Y	K	L	P	N	T	I	S	S	E	T	F	L	K	*

Positions that can be changed without affecting enzyme activity (demonstrated)
Positions -1 and -2 are not natural. The two amino acids are added due to cloning strategies

Positions 1 to 10: can be deleted
Position 36: G is tolerated
Position 40: M or V are tolerated
Position 41: S or N are tolerated
Position 43: A is tolerated
Position 46: V or N are tolerated
Position 91: A is tolerated
Positions 123 and 156: L are tolerated
Position 223: A and S are tolerated

Changes that affect enzyme activity (demonstrated)
Position 19: L to S
Position 38: I to S or N
Position 39: G to D or R
Position 40: L to Q
Position 42: L to R
Position 44: D to E G or H
Position 45: A to E or D
Position 46: Y to D
Position 47: I to R or N
Position 80: L to S
Position 144: D to E
Position 145: D to E
Position 146: G to E
Position 147: G to S

FIG. 5

Group I Intron Encoded Endonucleases and Related Endonucleases

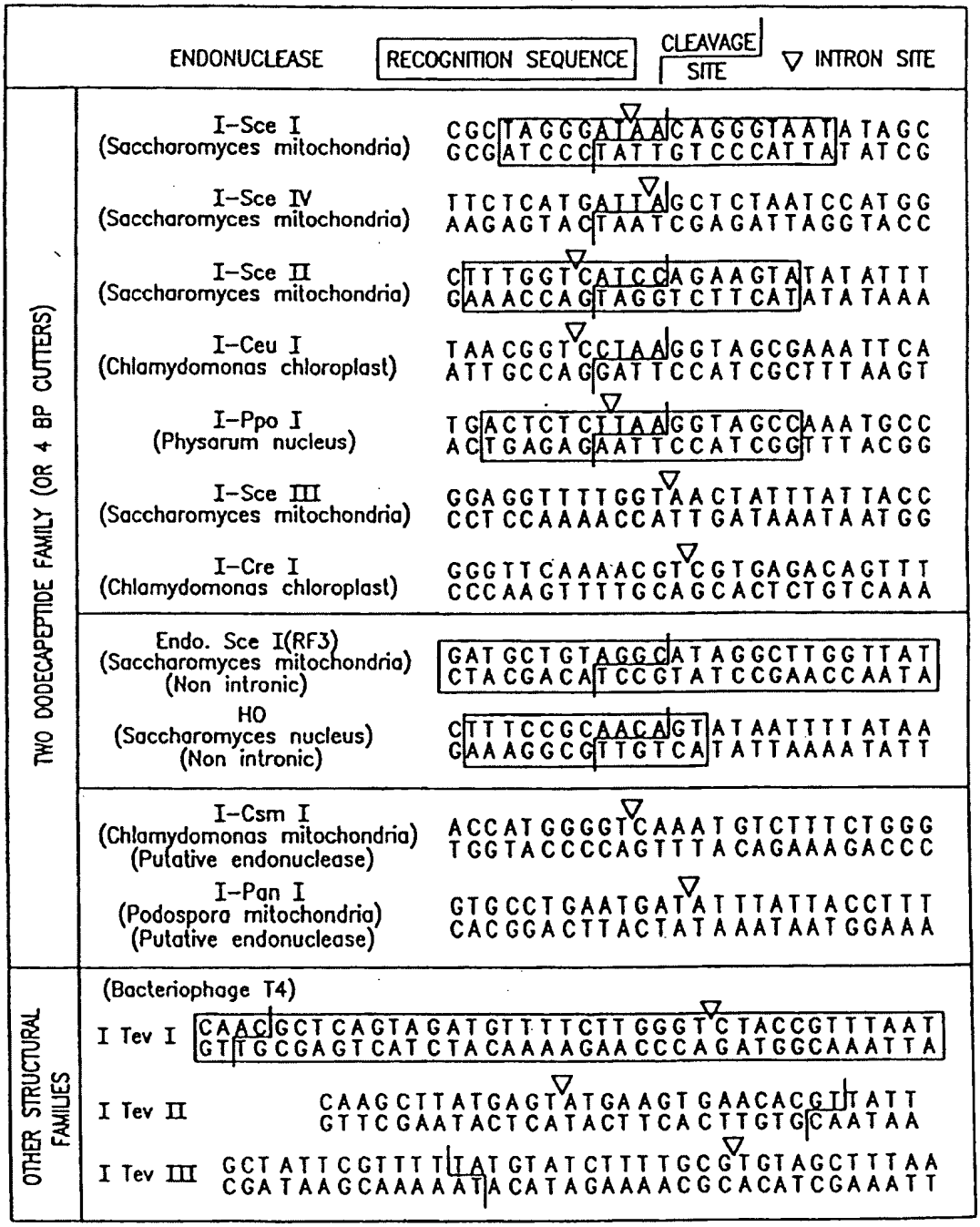


FIG. 6

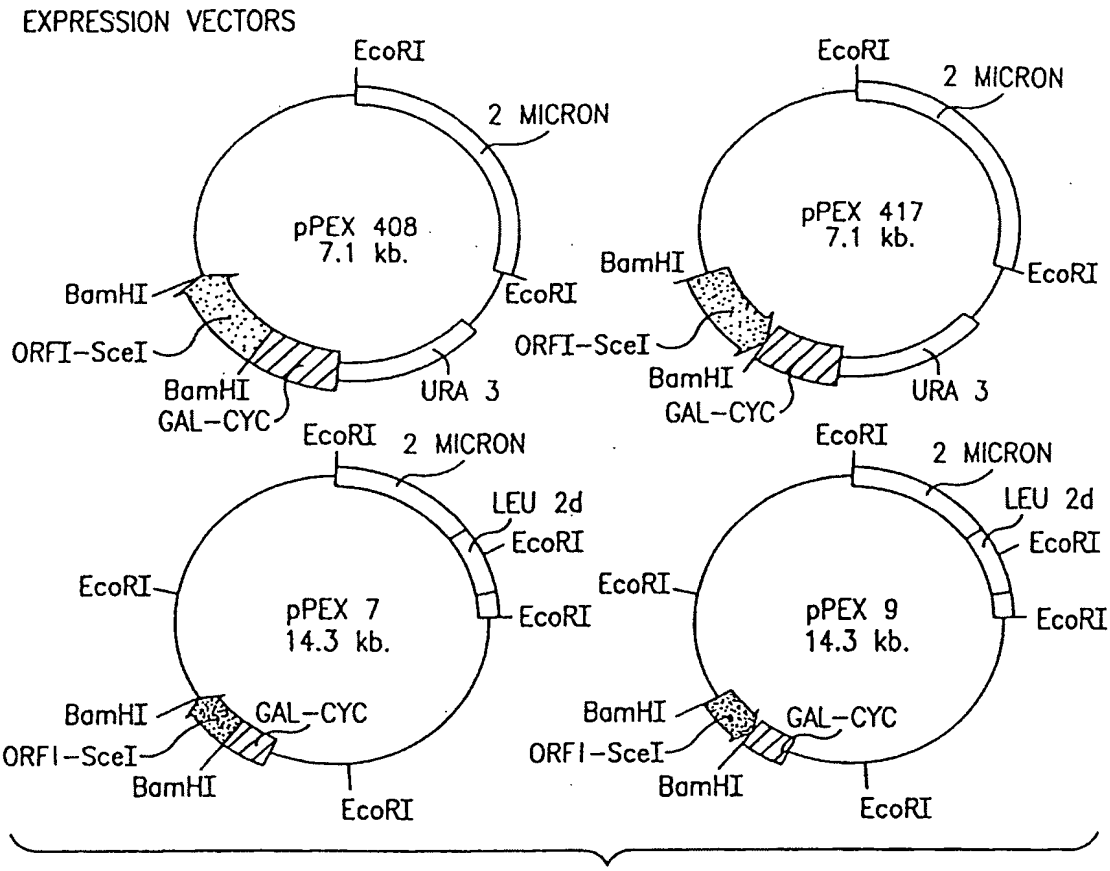


FIG. 7

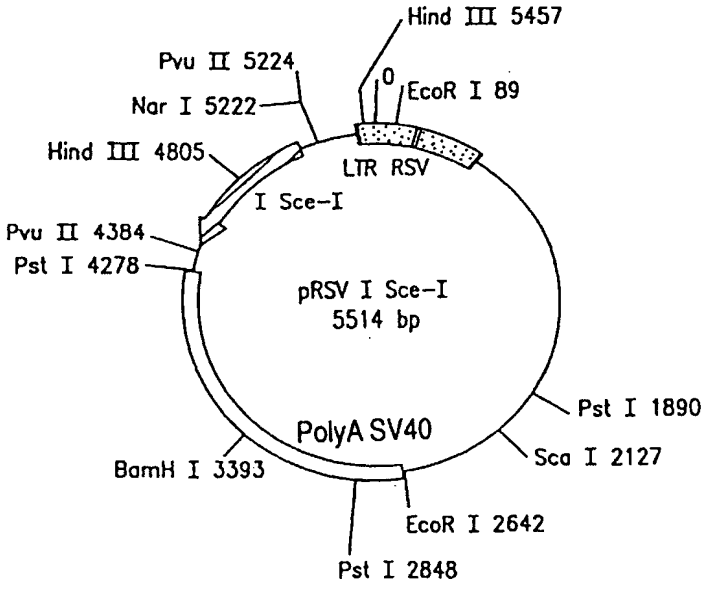


FIG. 8

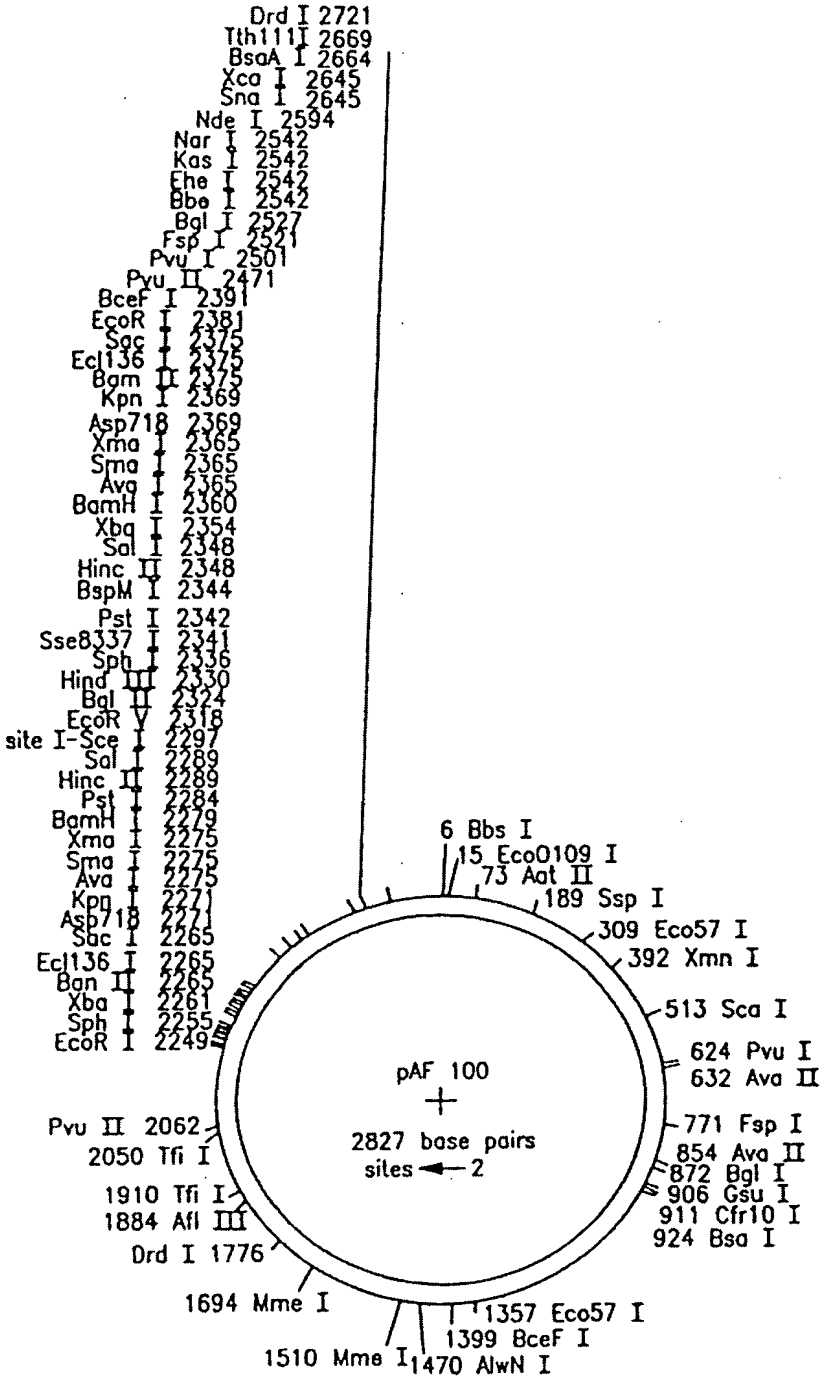


FIG. 9

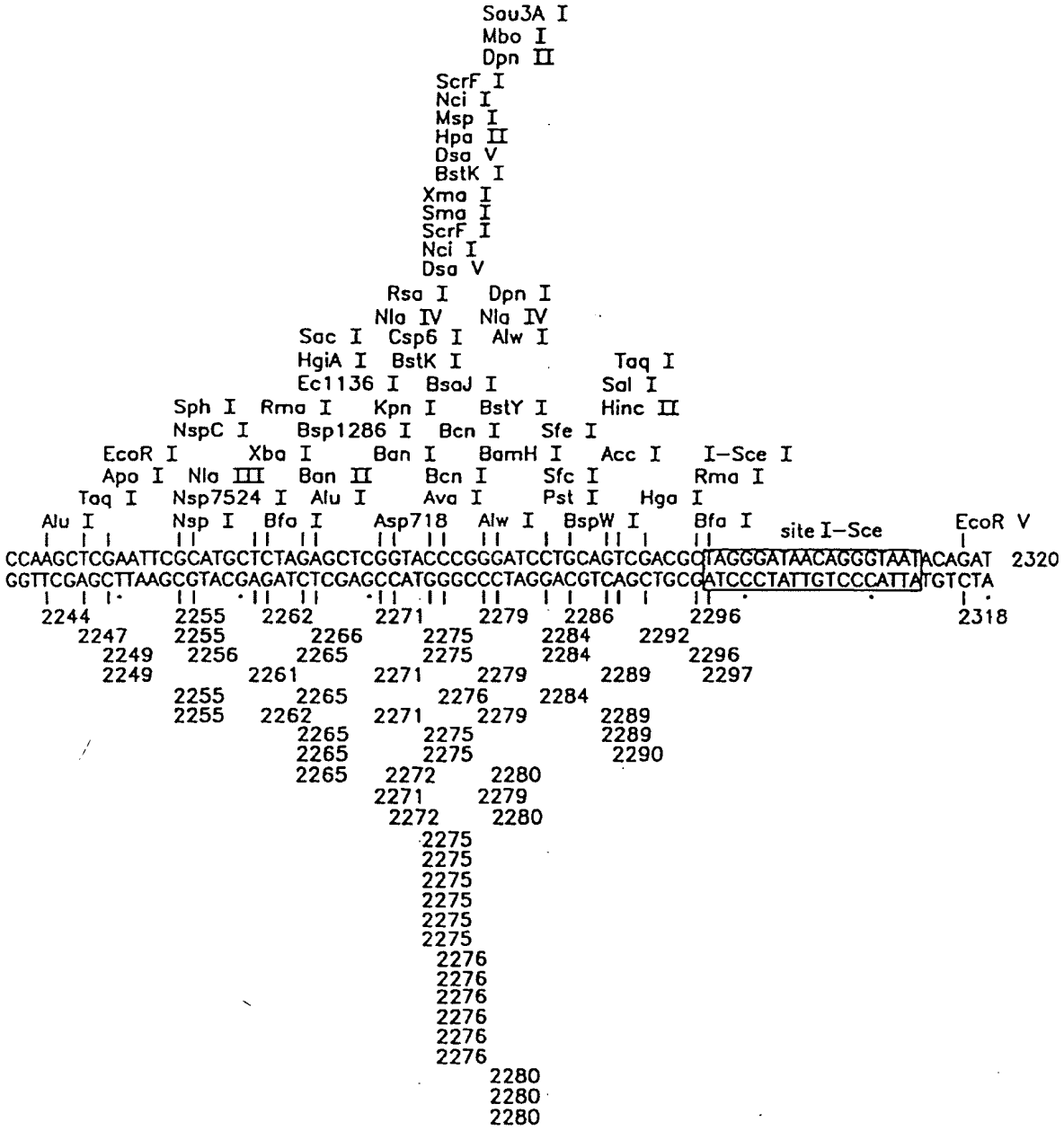


FIG. 10A

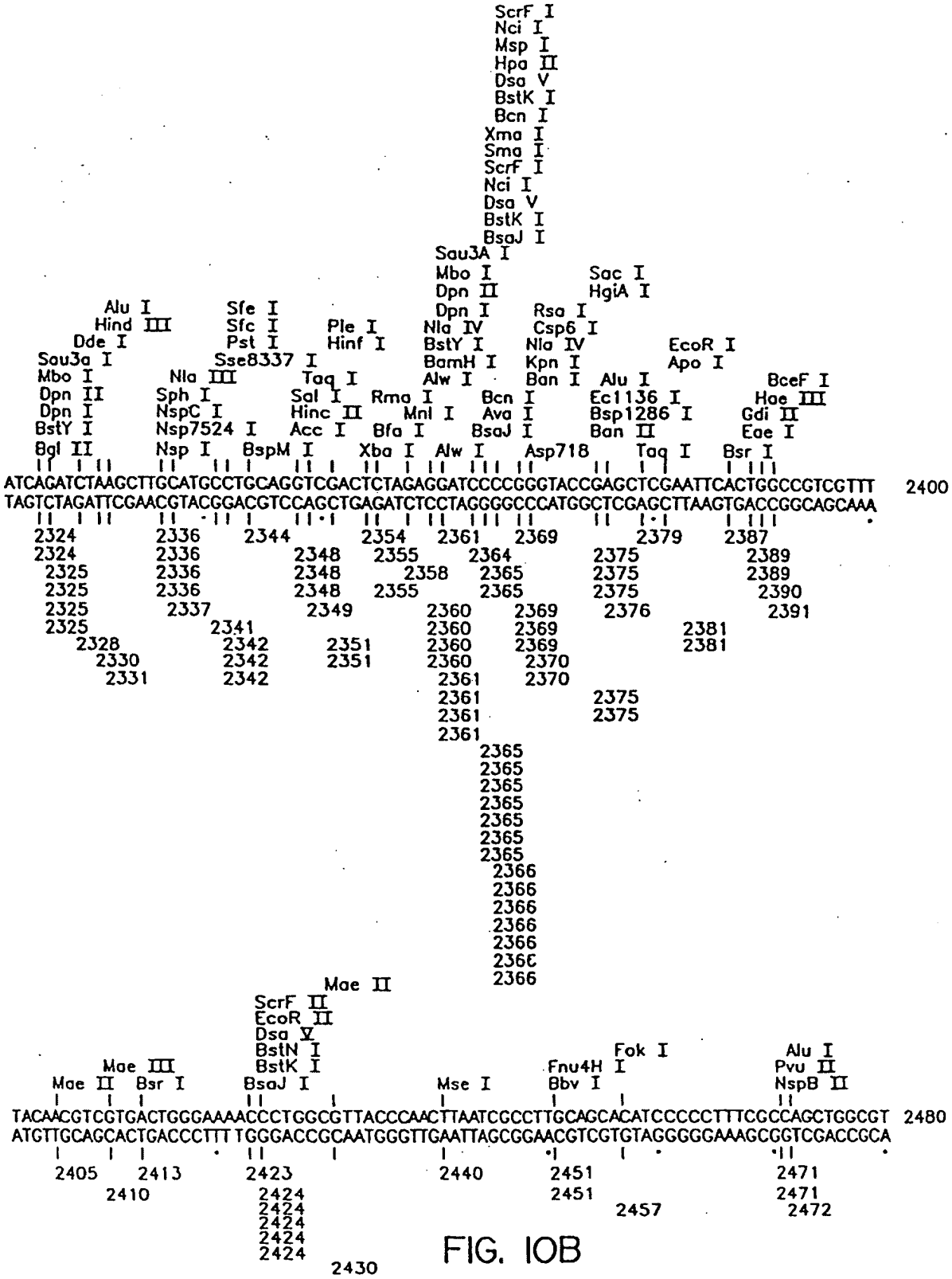
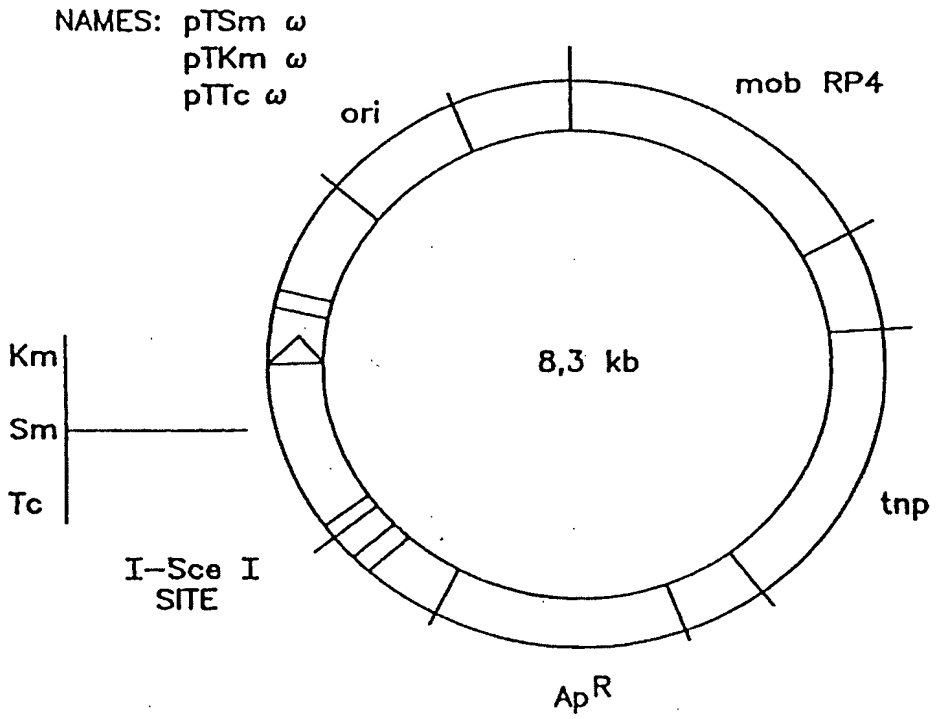
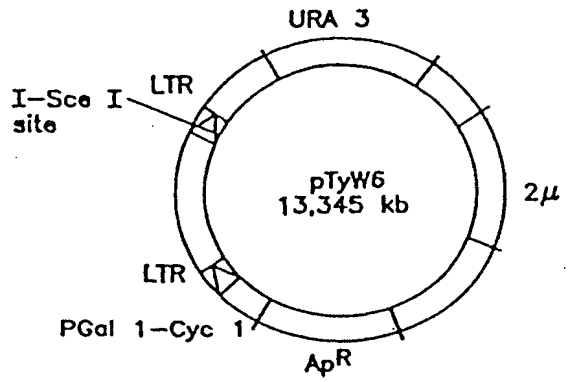


FIG. 10B



Construction: pGP 704 from De Lorenzo, with transposase gene and insertion of the linker[I-SceI] in NotI unique site

FIG. 11



Construction: pD 123, from J.D. Boeke
with insertion of a linker[I–SceI–NotI] in BamHI

FIG. 12

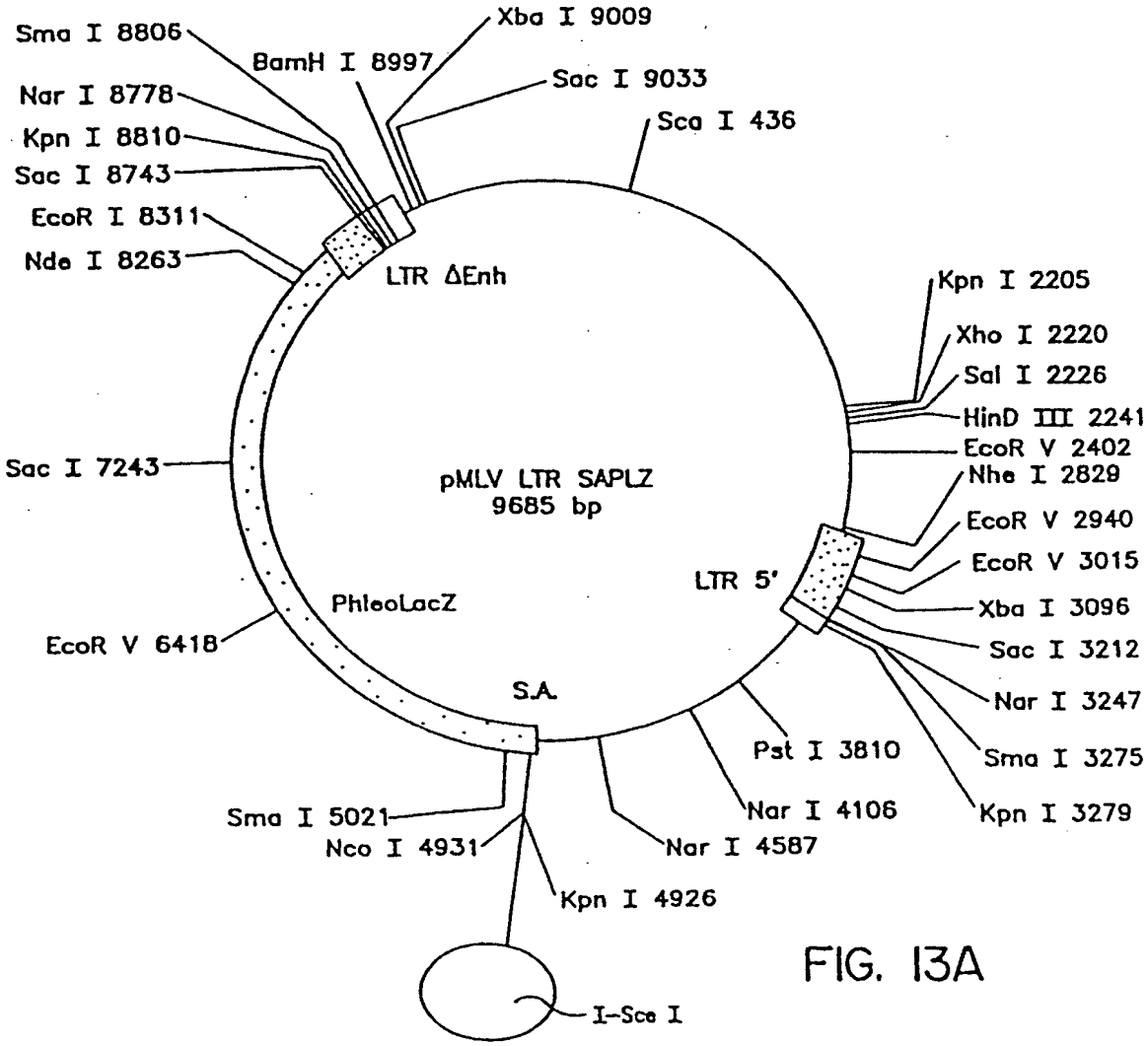


FIG. 13A

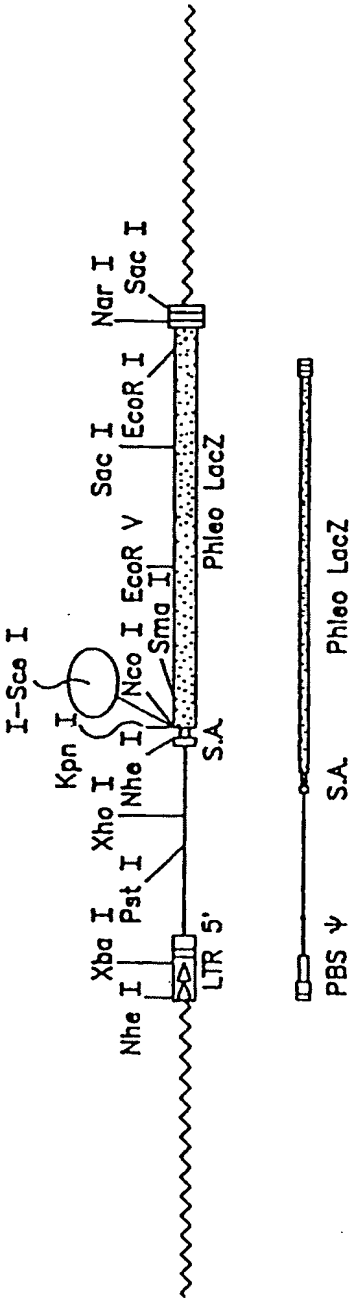


FIG. 13B

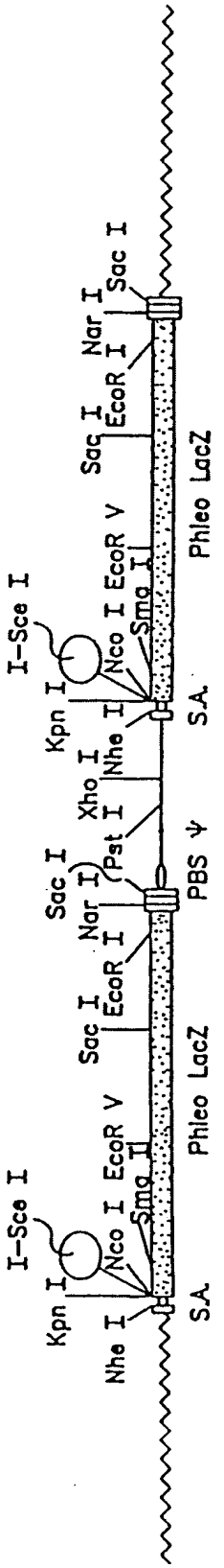


FIG. 13C

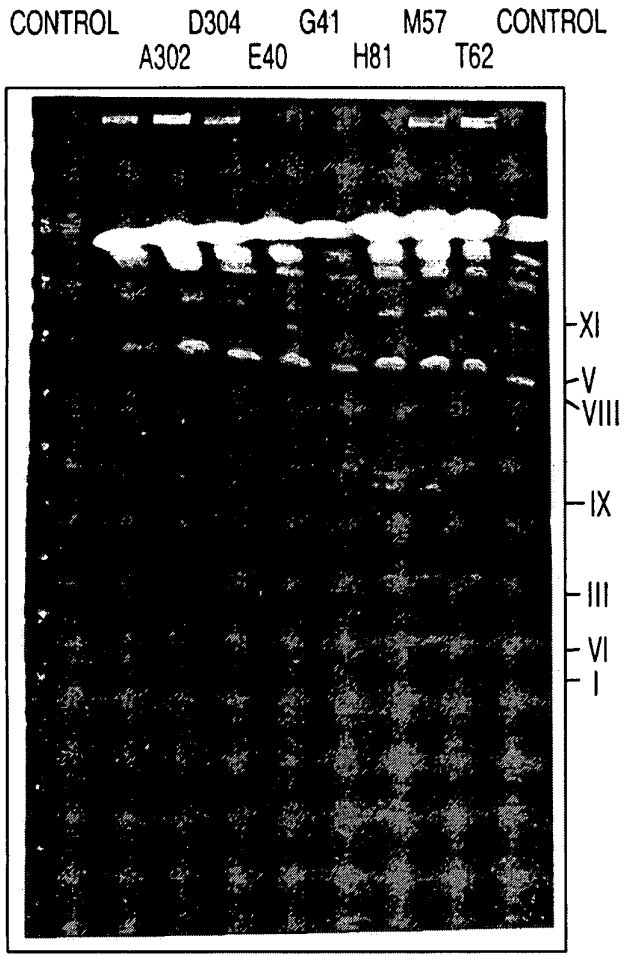


FIG. 14A

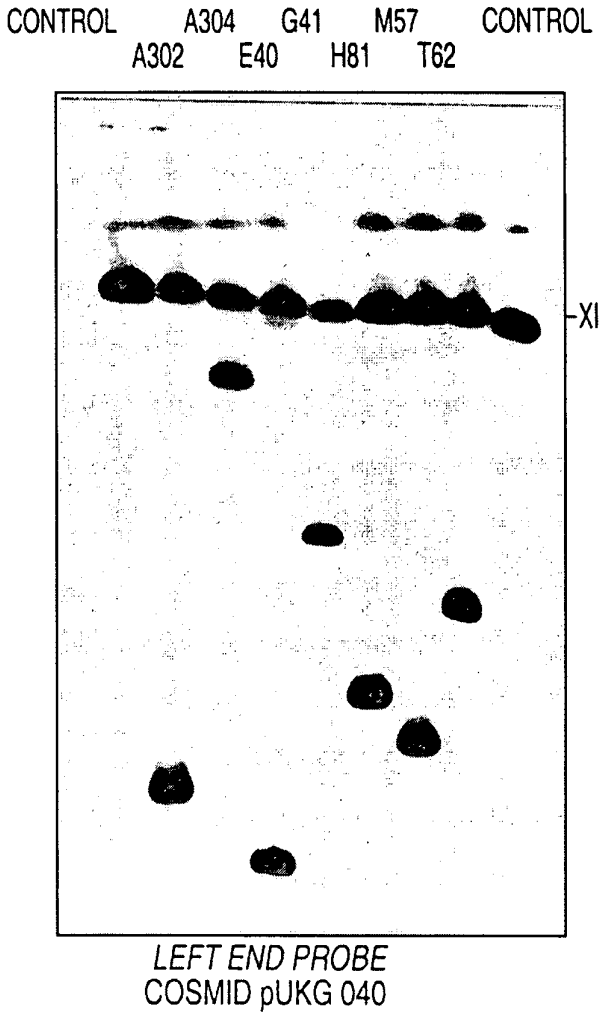


FIG. 14B

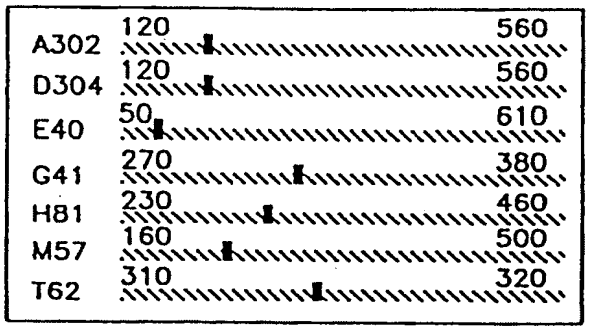


FIG. 15A

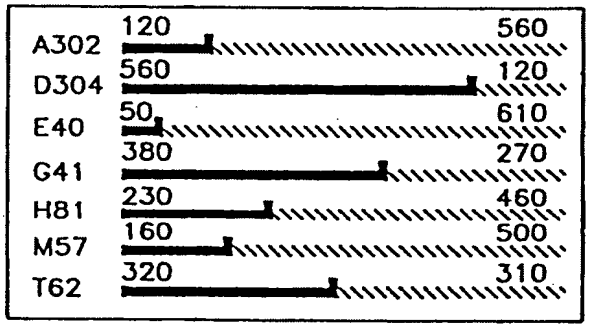


FIG. 15B

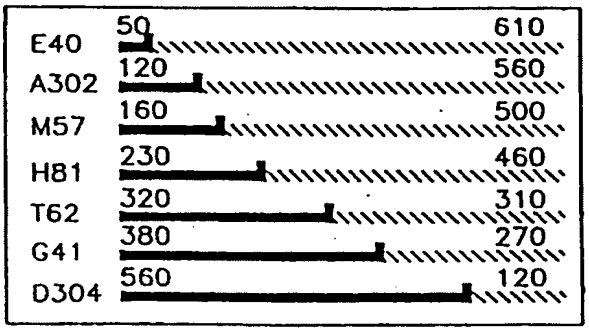


FIG. 15C

FIG. 15D

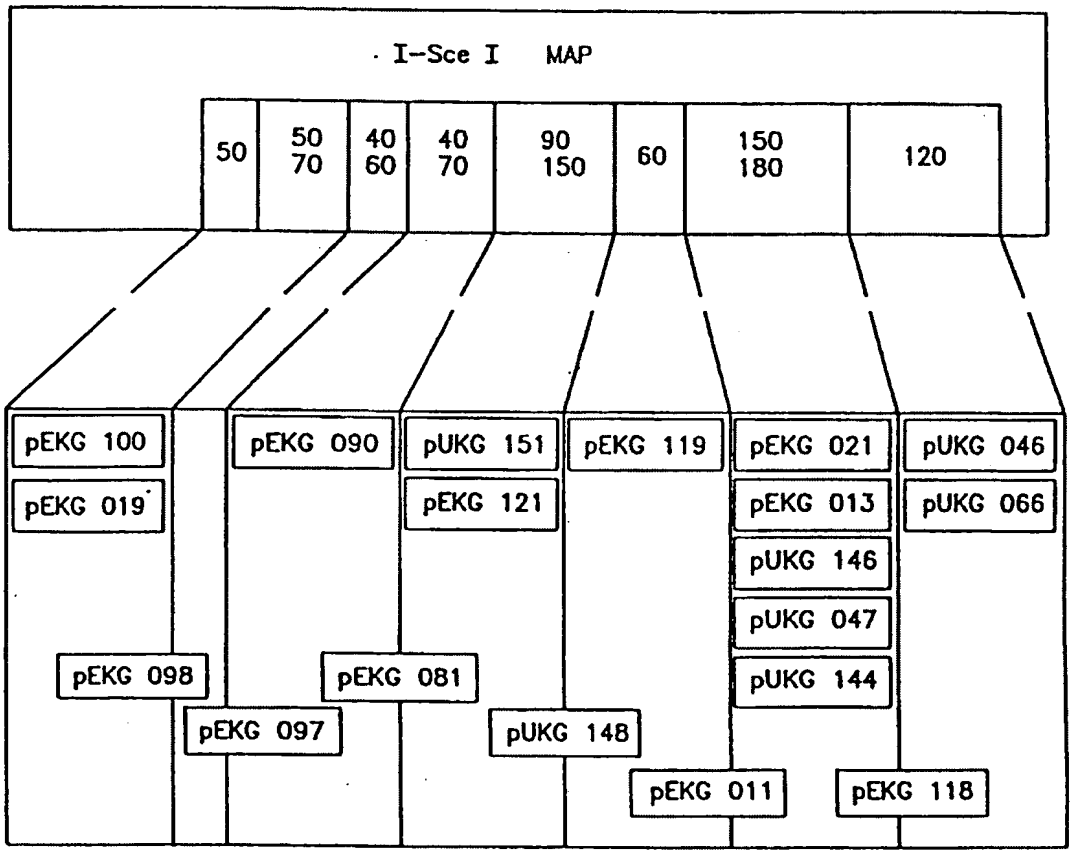


FIG. 15E

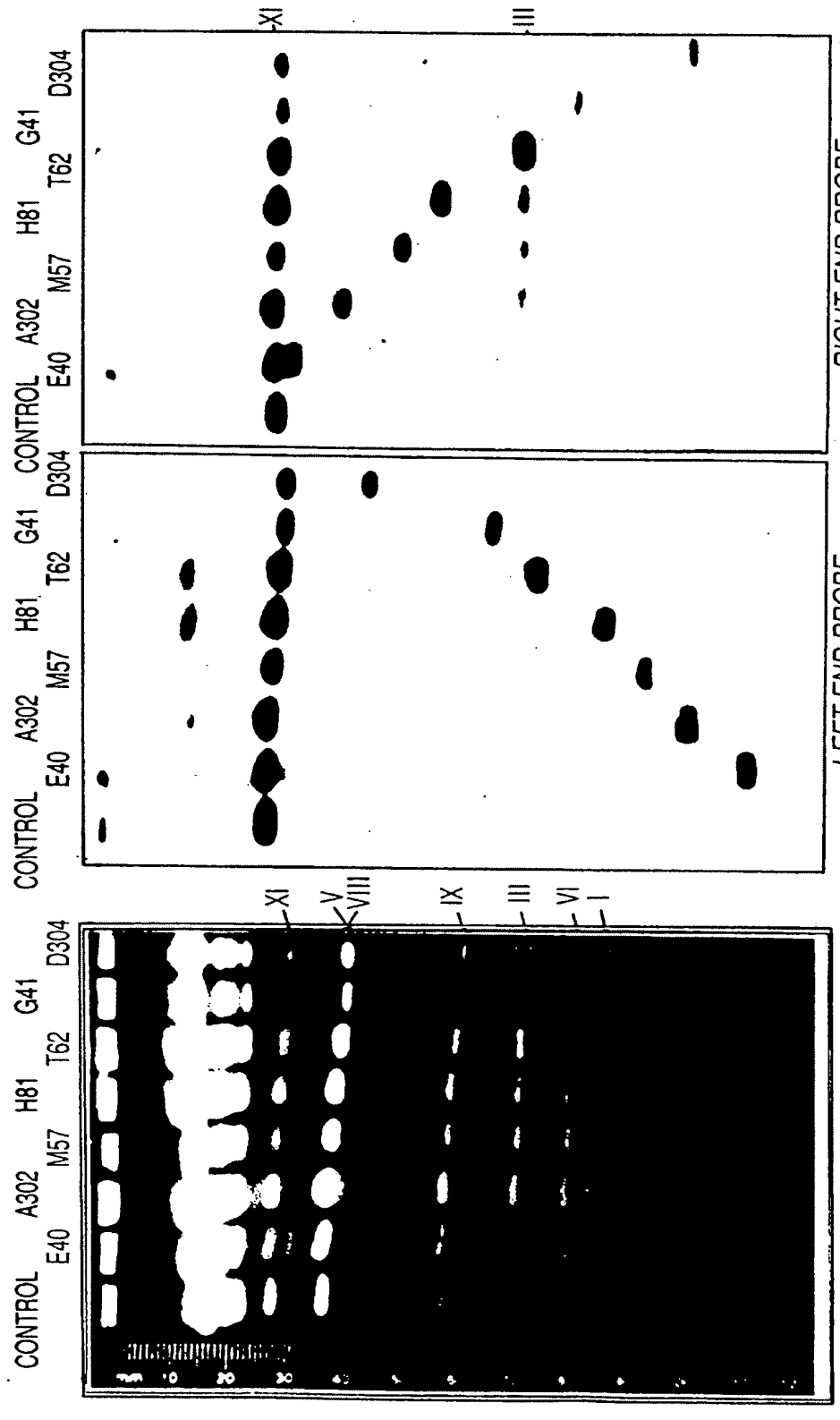


FIG. 16A

FIG. 16B

FIG. 16C

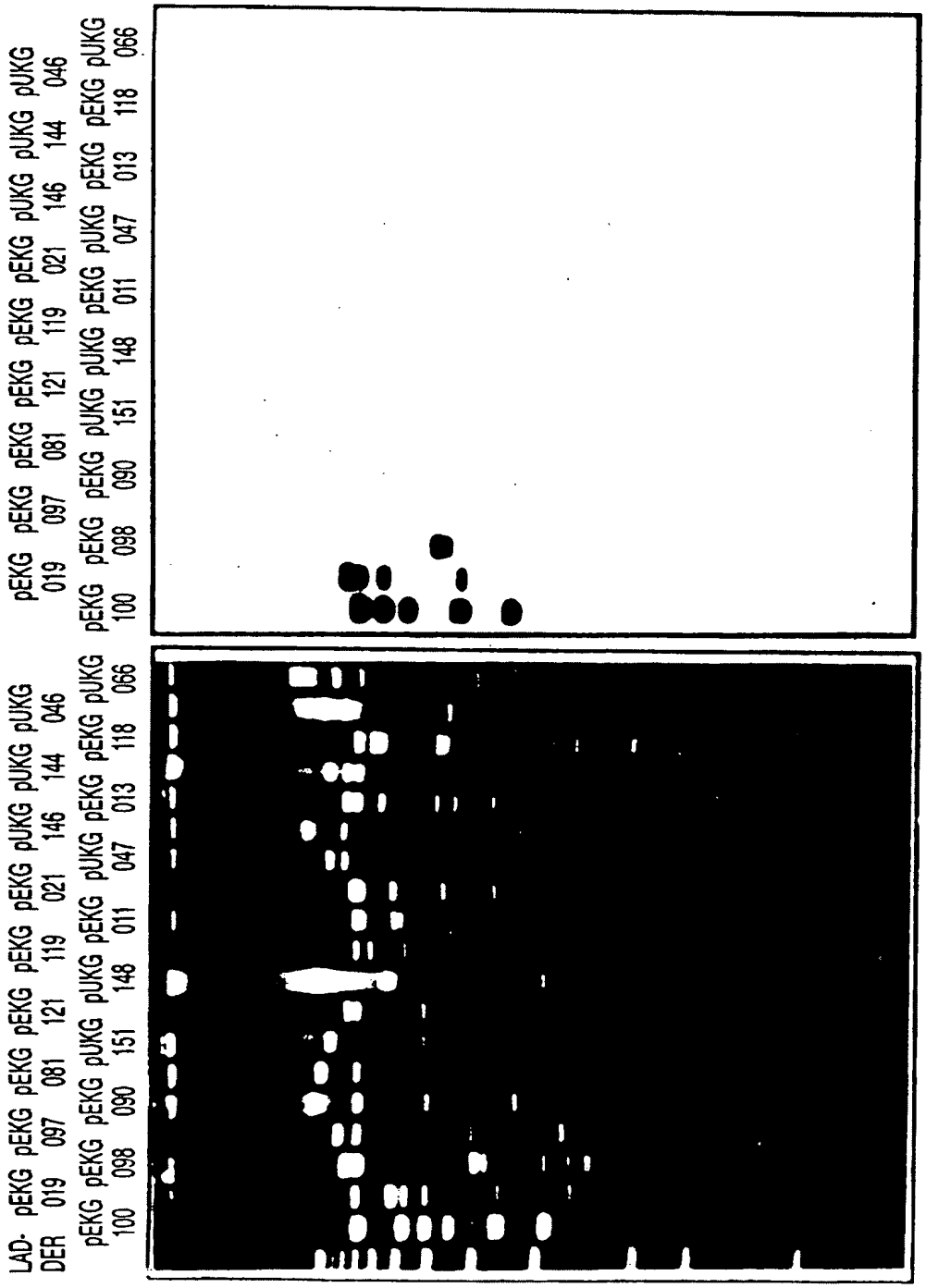


FIG. 17A

FIG. 17B

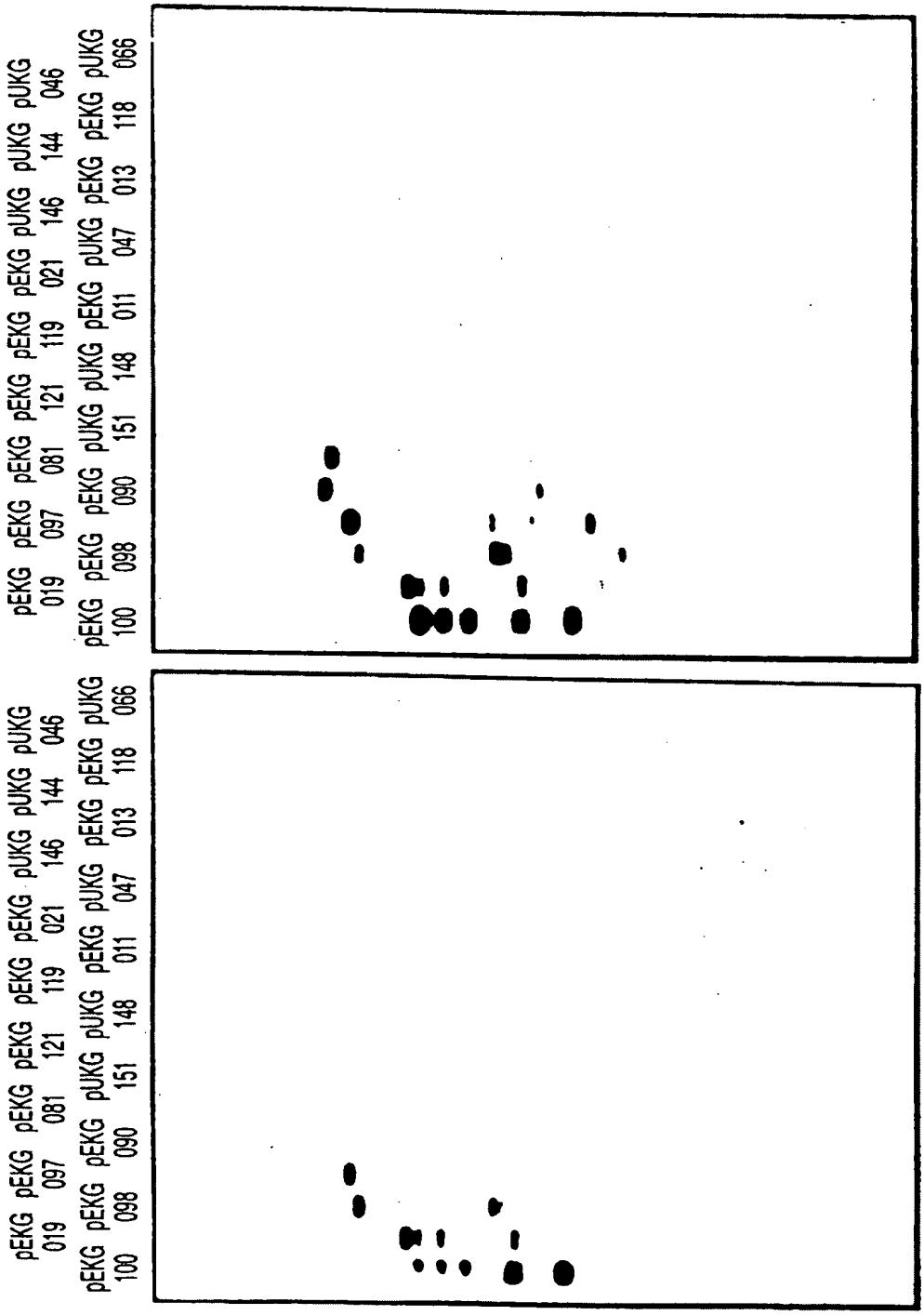


FIG. 17C

FIG. 17D

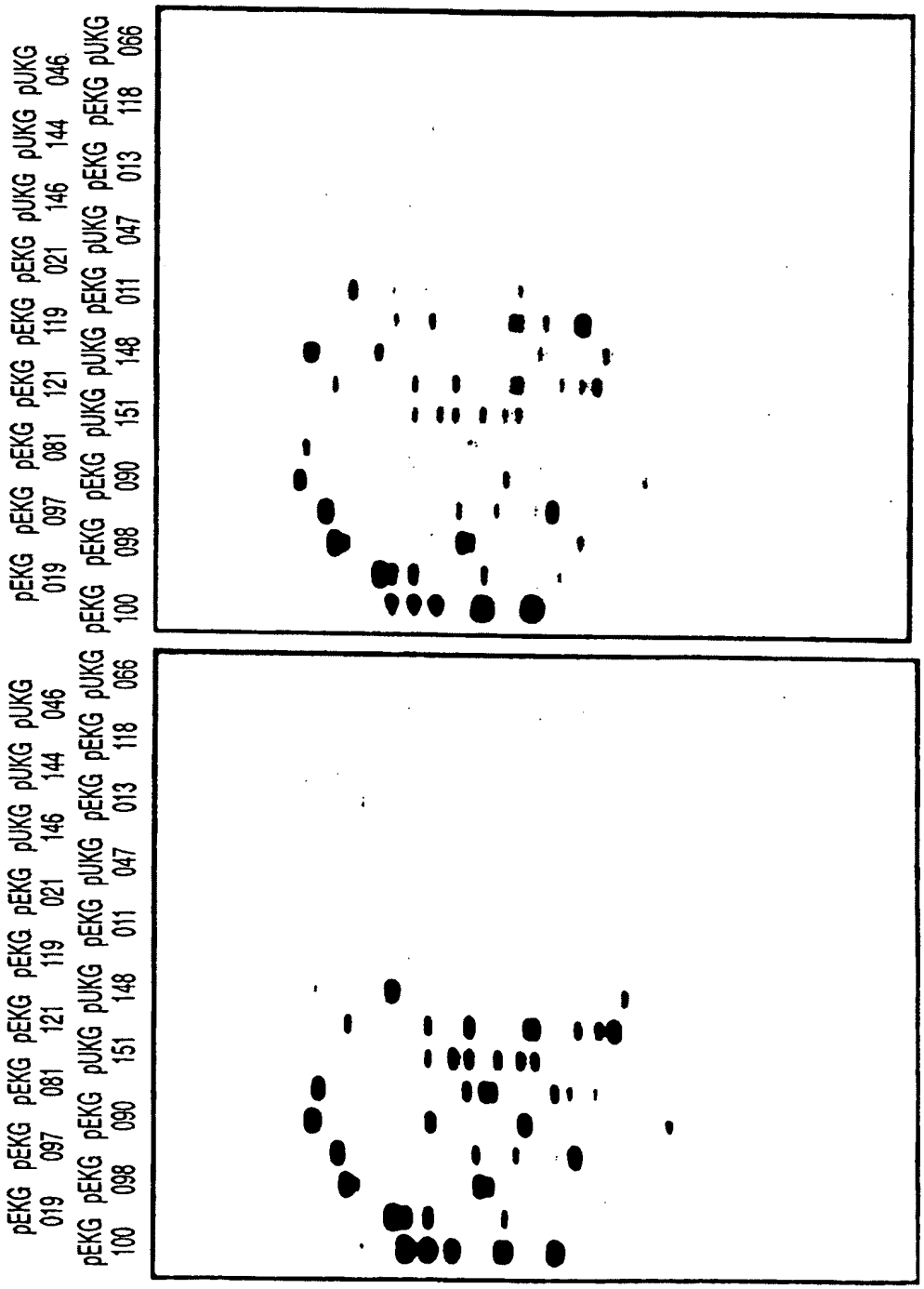


FIG. 17E

FIG. 17F

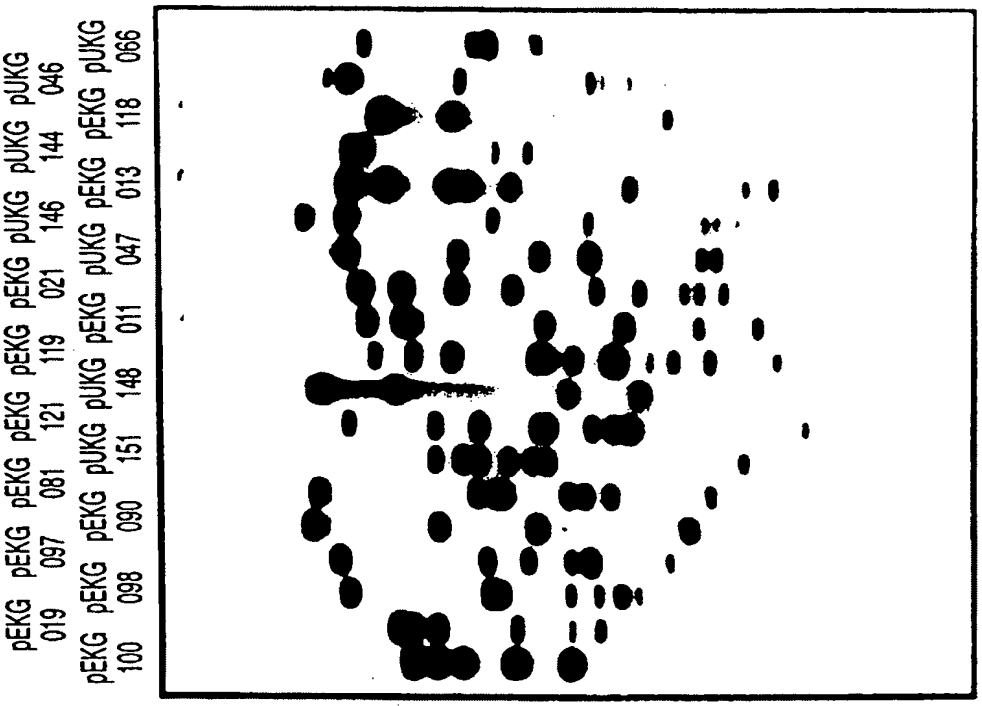


FIG. 17G

FIG. 17H

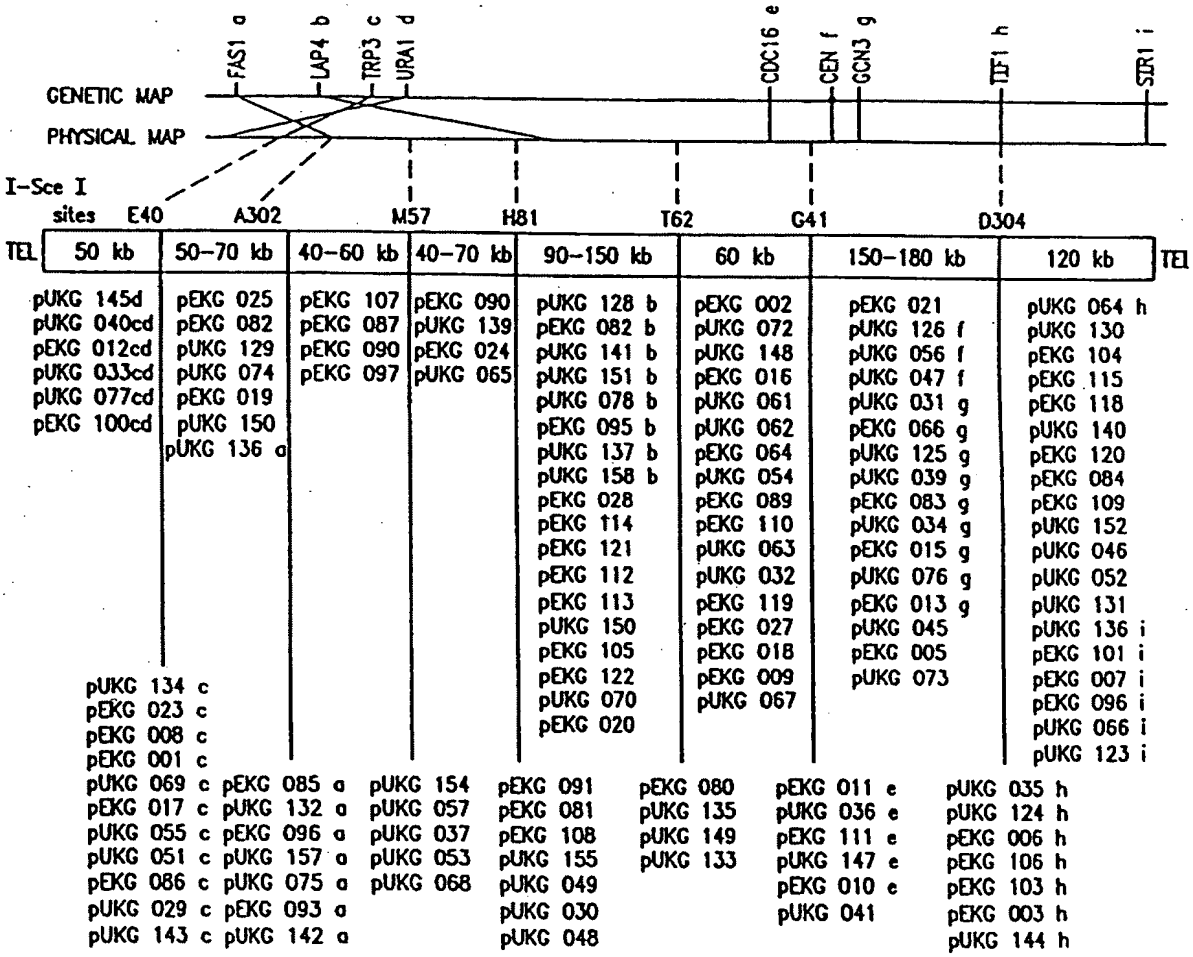


FIG. 18

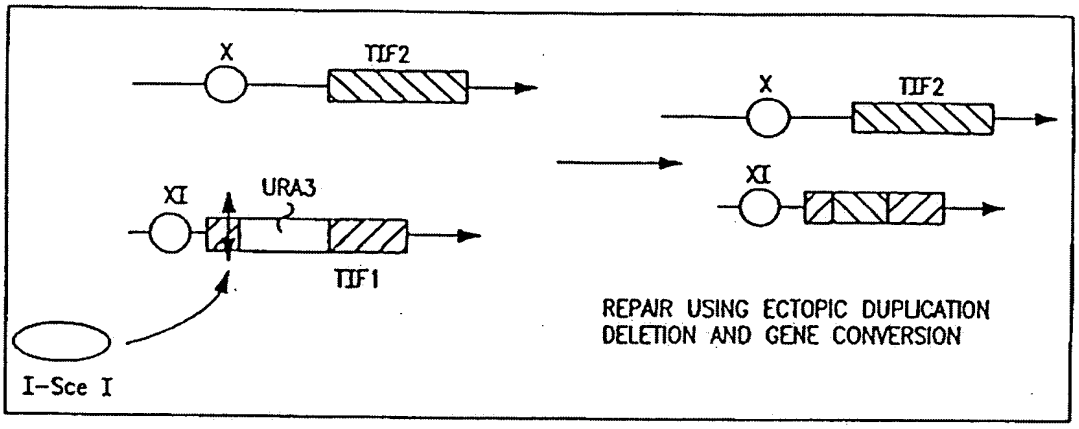


FIG. 19A

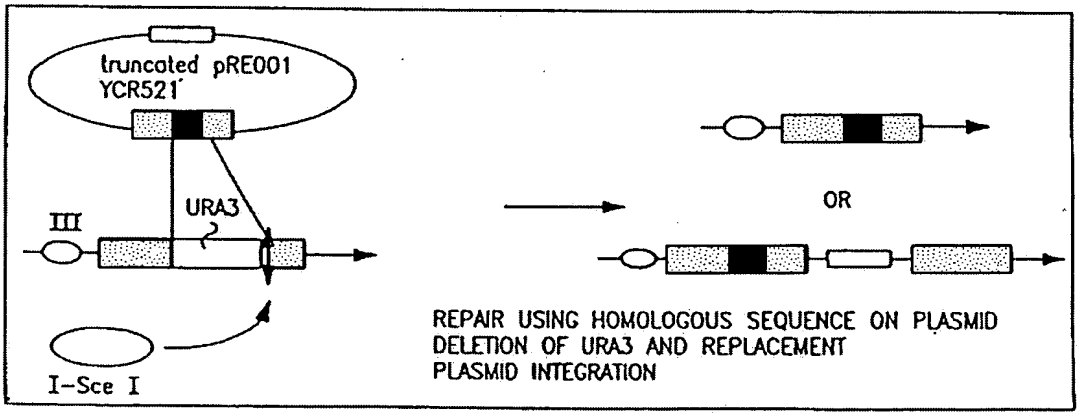
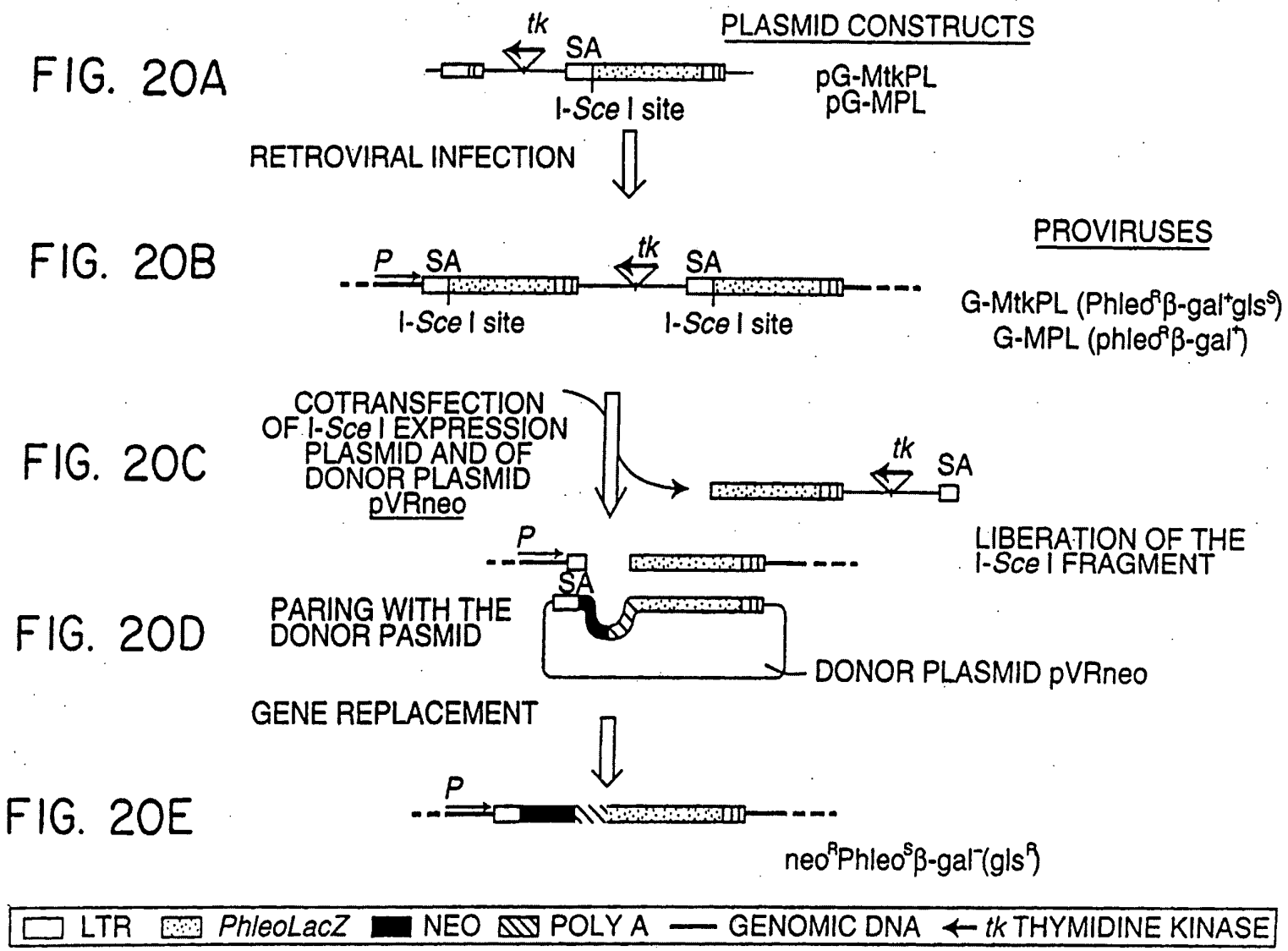


FIG. 19B



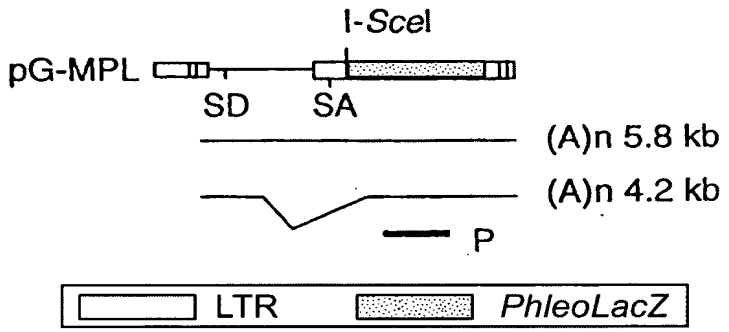


FIG. 2IA

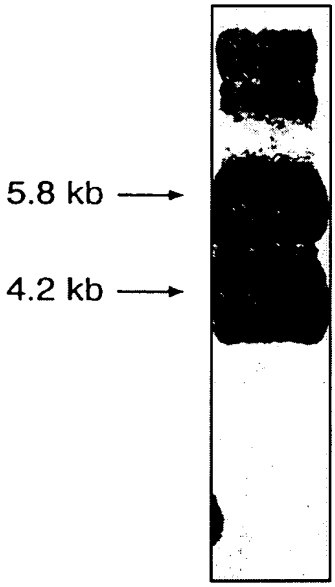


FIG. 21B

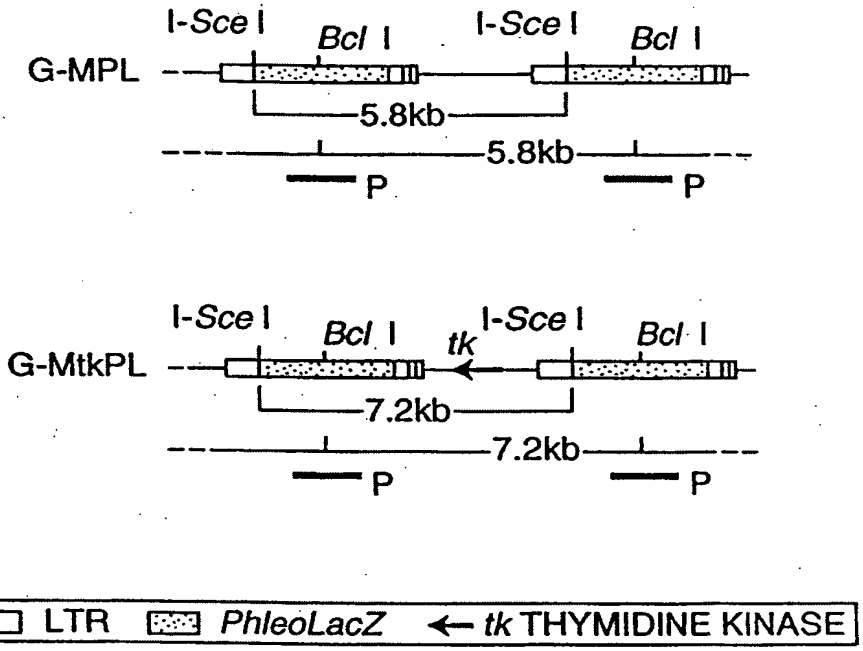


FIG. 22A

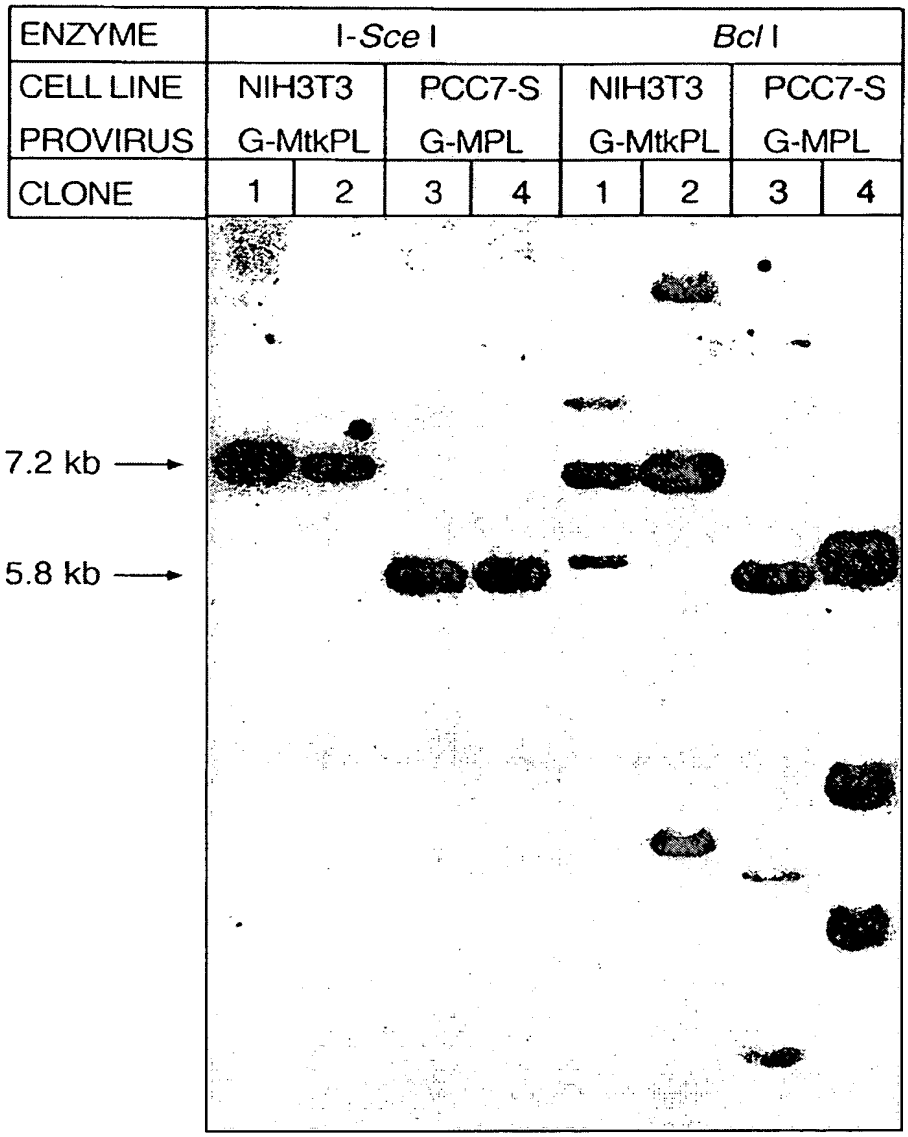


FIG. 22B

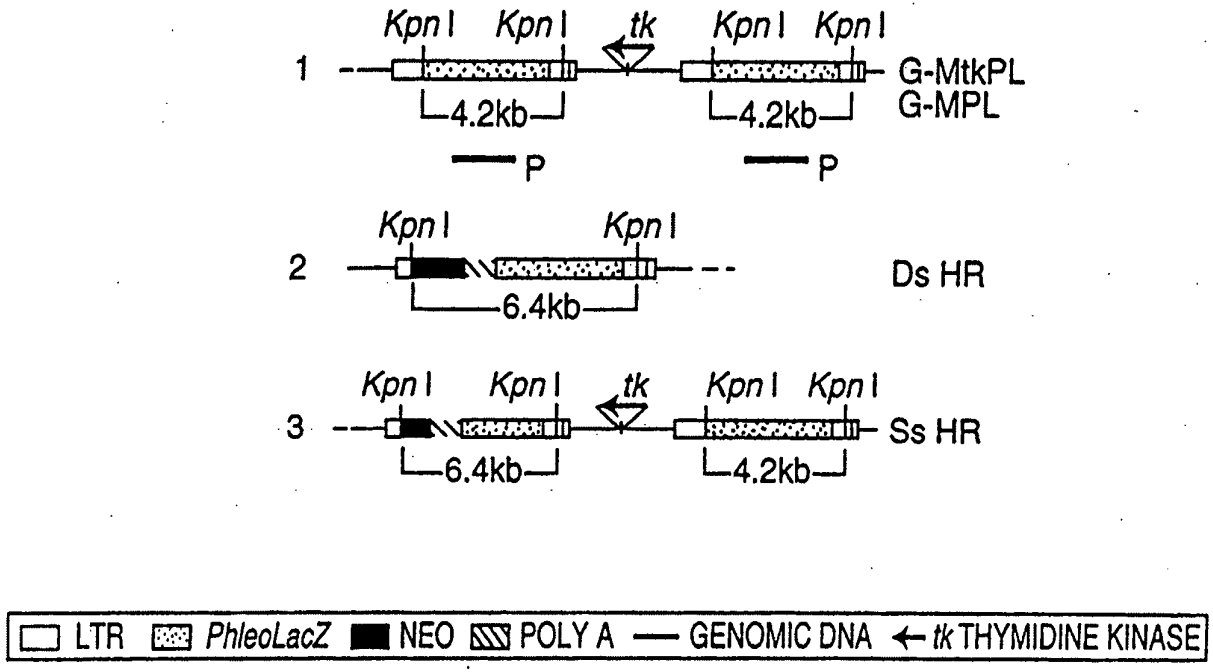


FIG. 23A

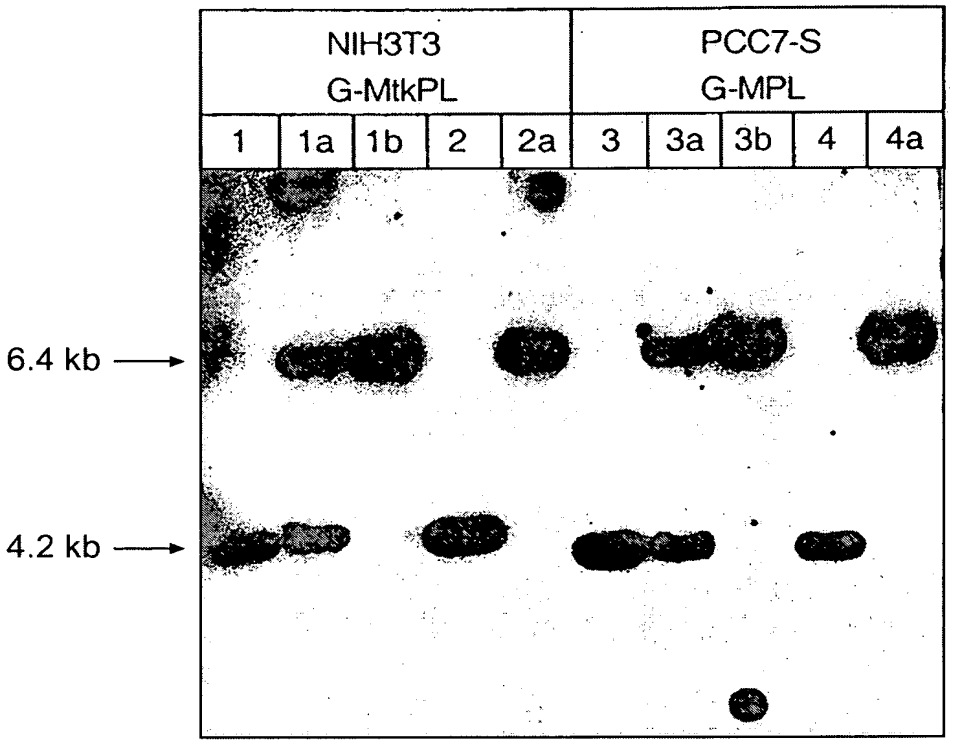


FIG. 23B

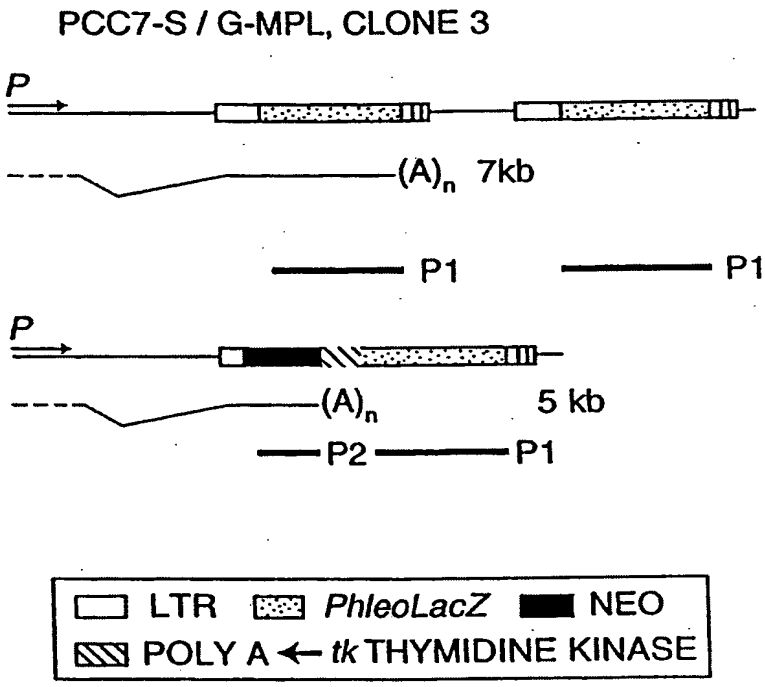


FIG. 24A

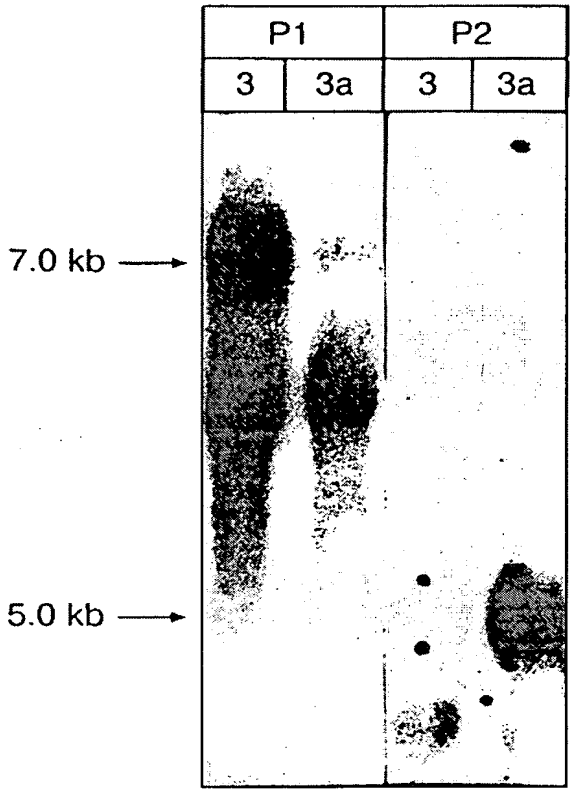


FIG. 24B

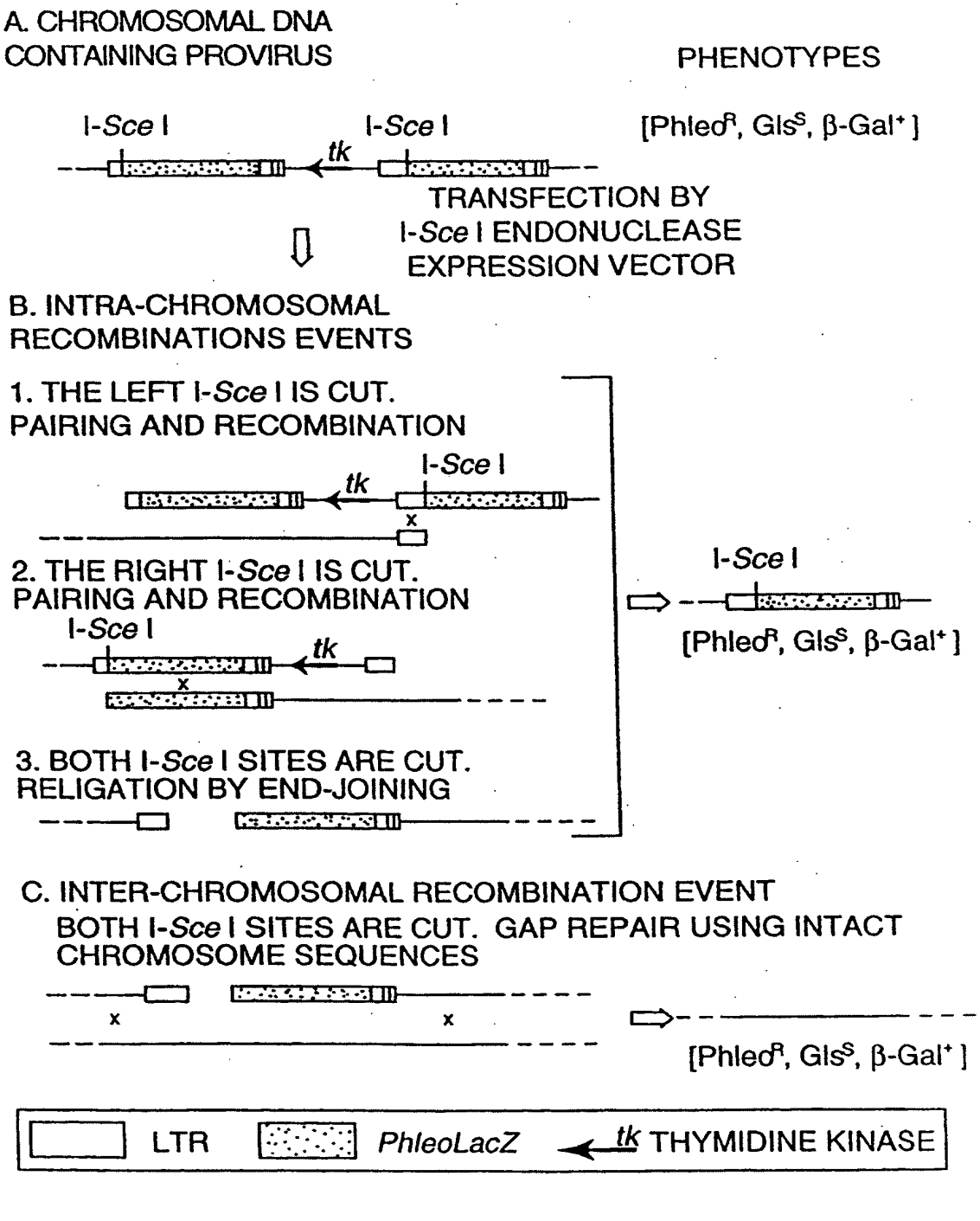


FIG. 25

A. PARENTAL DNA, G-MtkPL

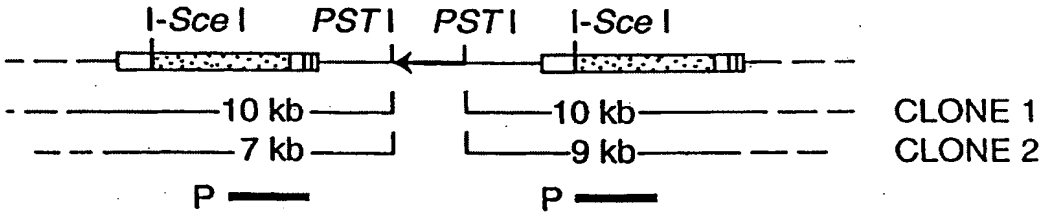


FIG. 26A

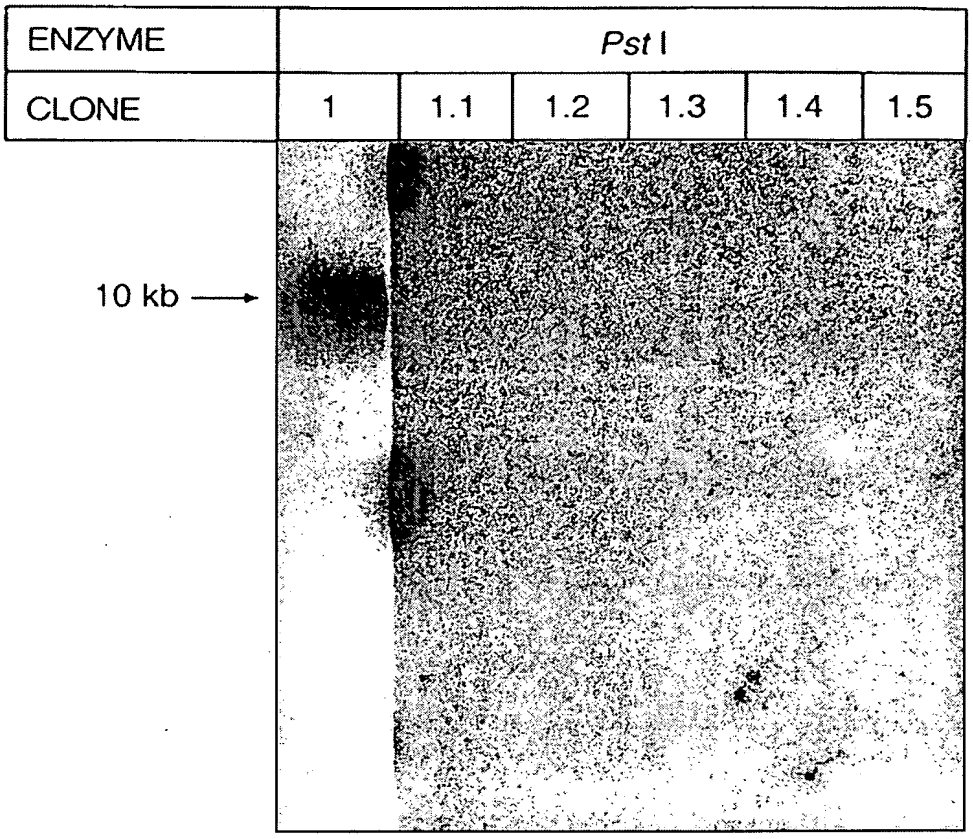


FIG. 26B

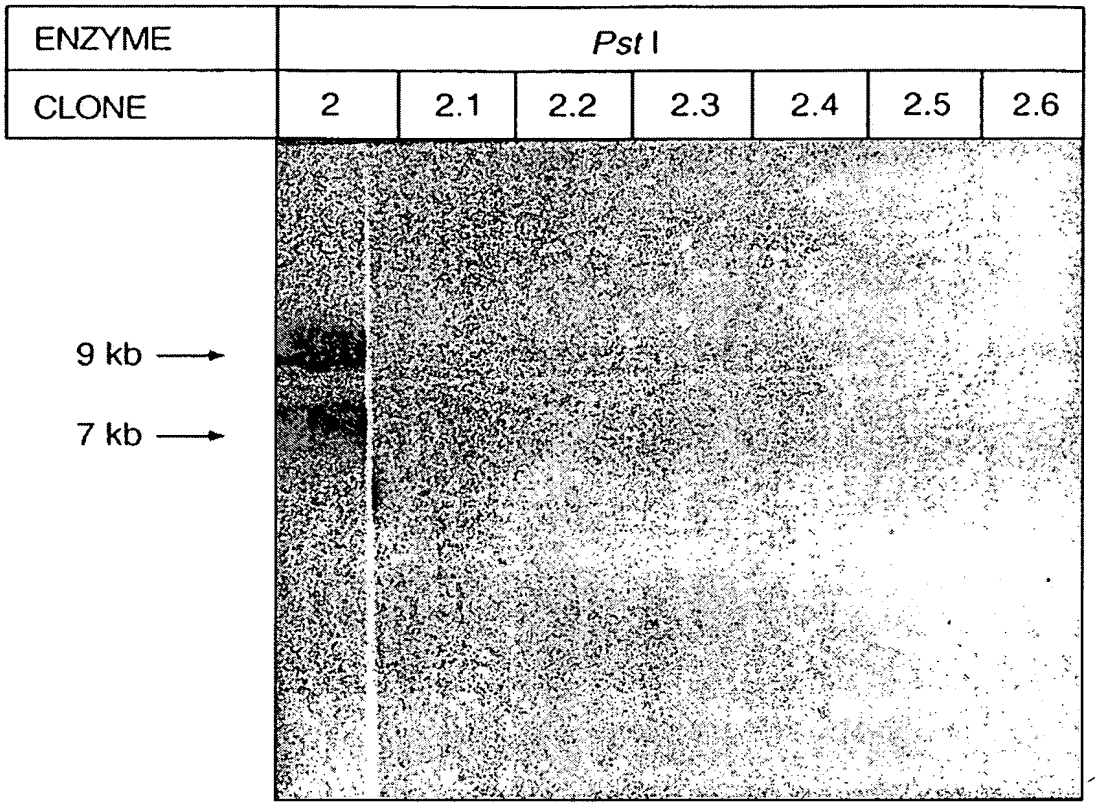
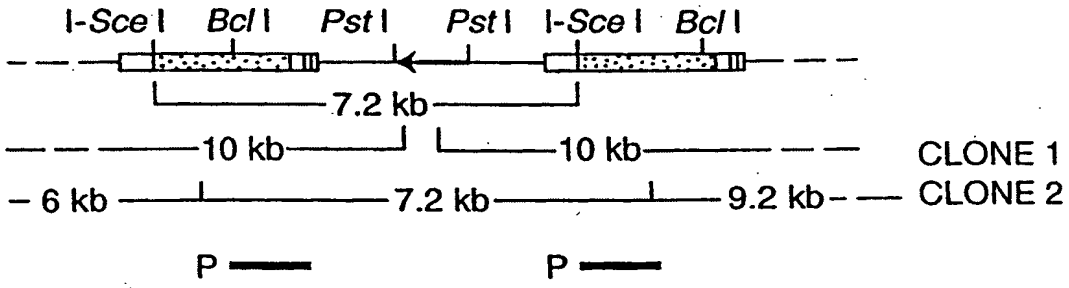


FIG. 26C

1. PARENTAL DNA, G-MtkPL



2. INTRA-MOLECULAR RECOMBINATION EVENT

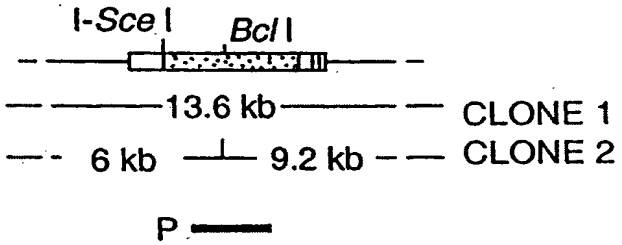


FIG. 27A

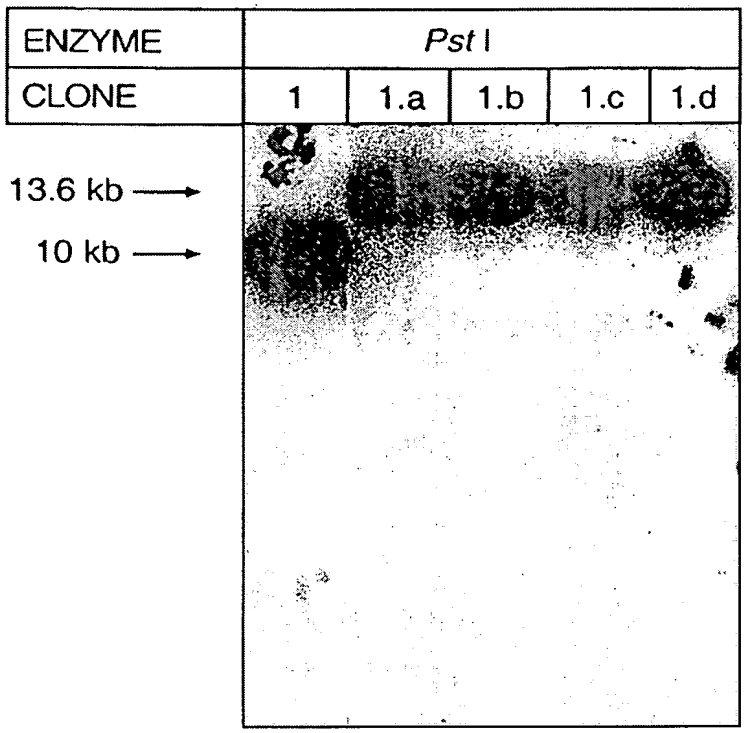


FIG. 27B

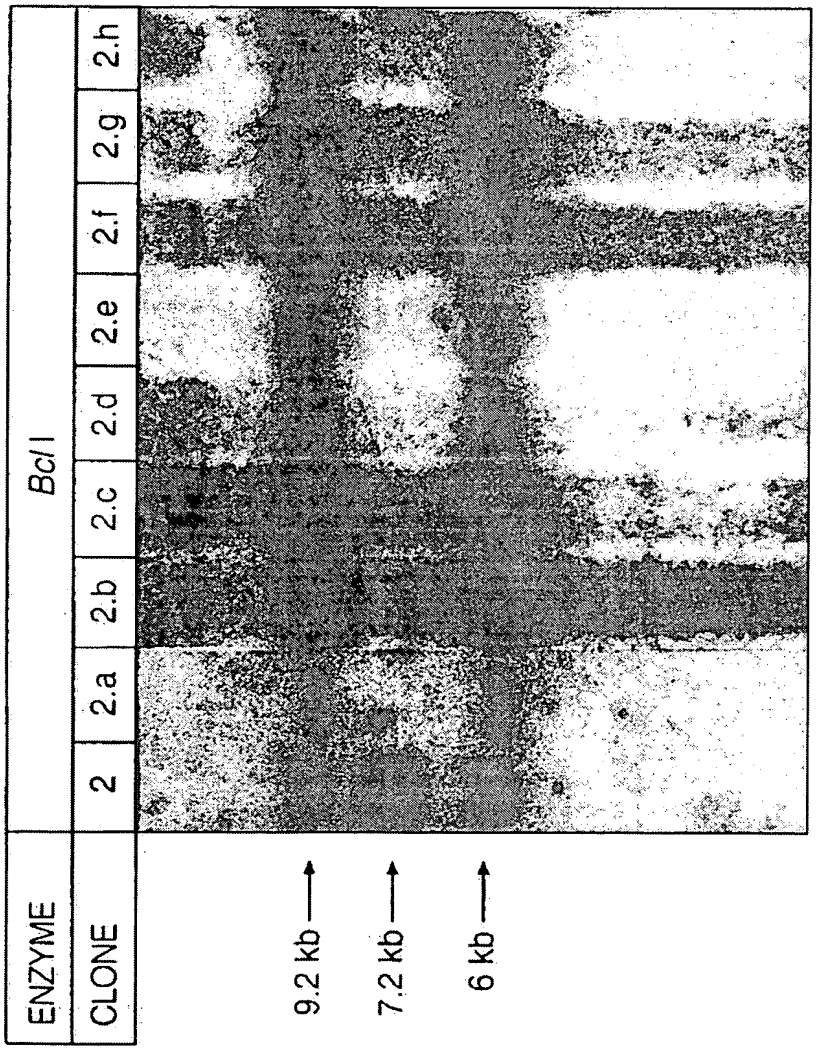


FIG. 27C

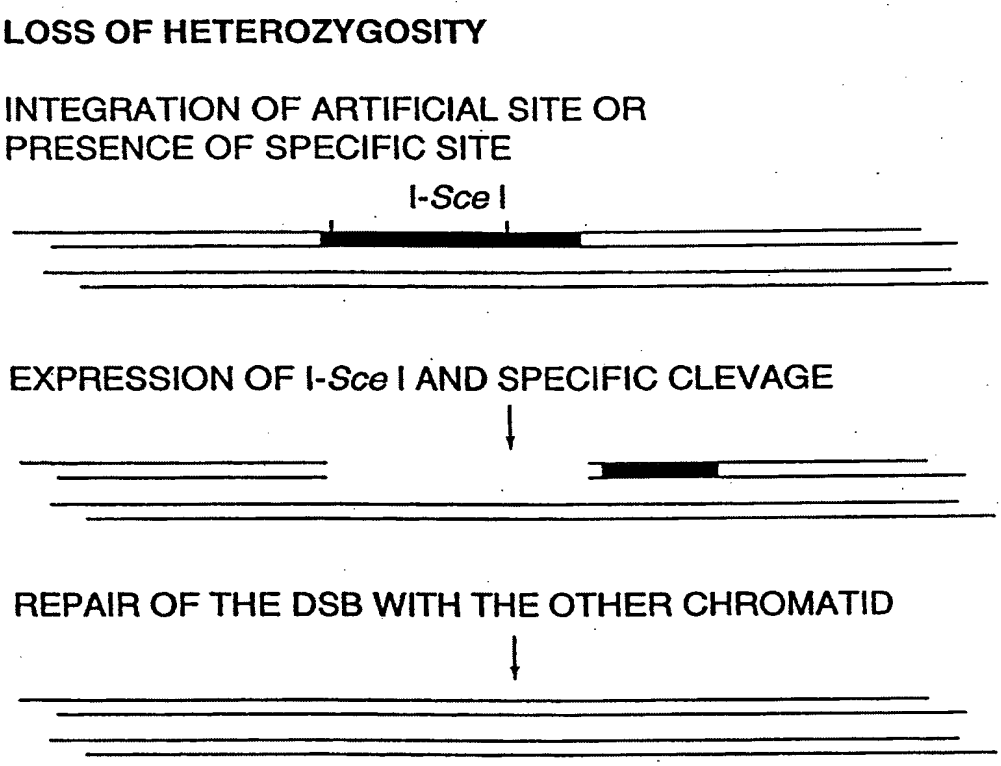


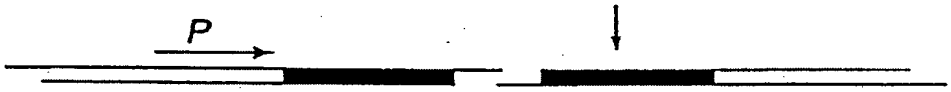
FIG. 28

CONDITIONAL ACTIVATION (TANDEM REPEAT)

INTEGRATION OF ARTIFICIAL SITE BETWEEN TANDEM REPEATS

GENE X INACTIVE

EXPRESSION OF I-*Sce* I AND SPECIFIC CLEAVAGE



REPAIR OF THE DSB BY SINGLE STRAND ANNEALING

GENE X ACTIVE

FIG. 29

ONE STEP REARRANGEMENT

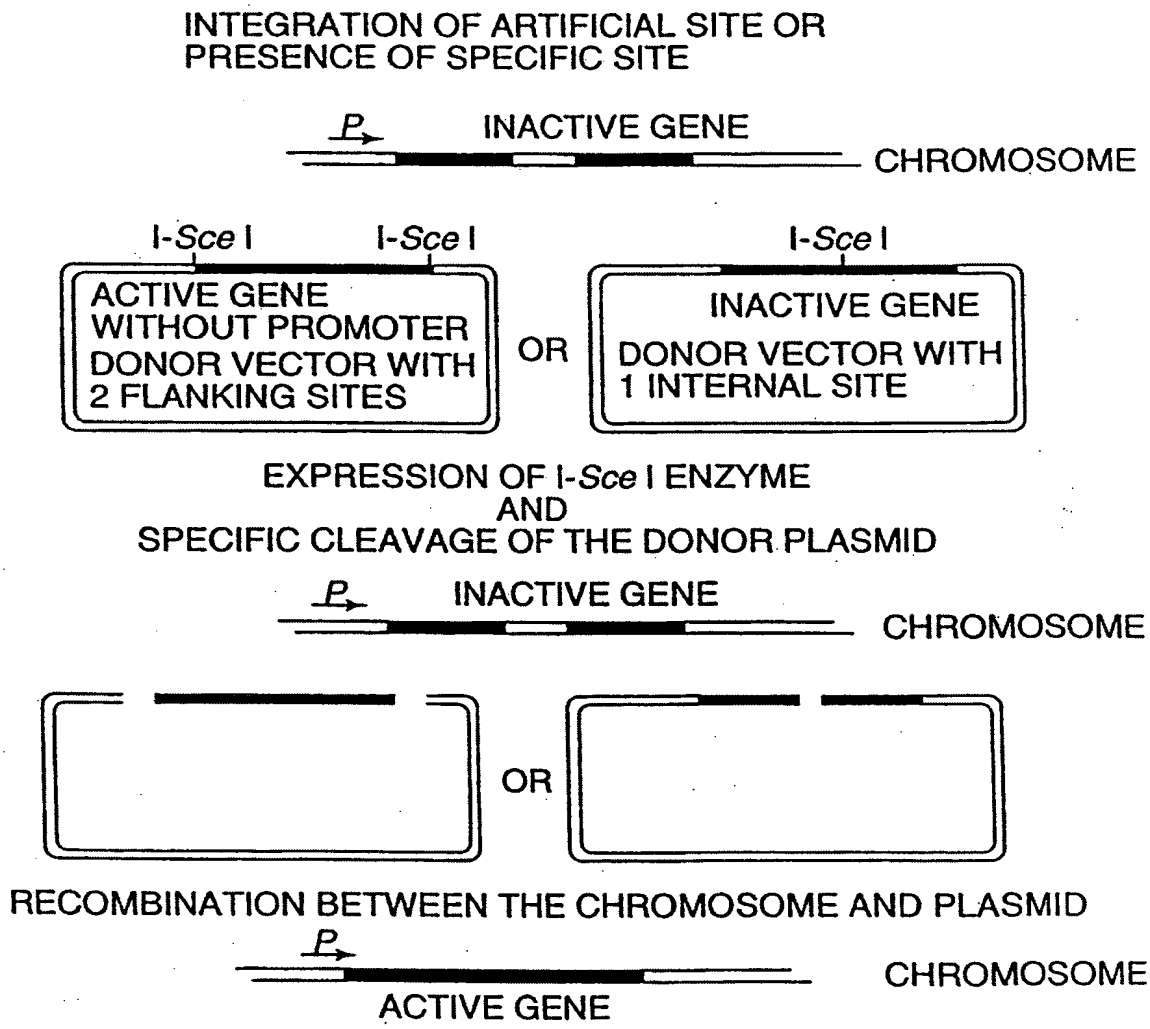
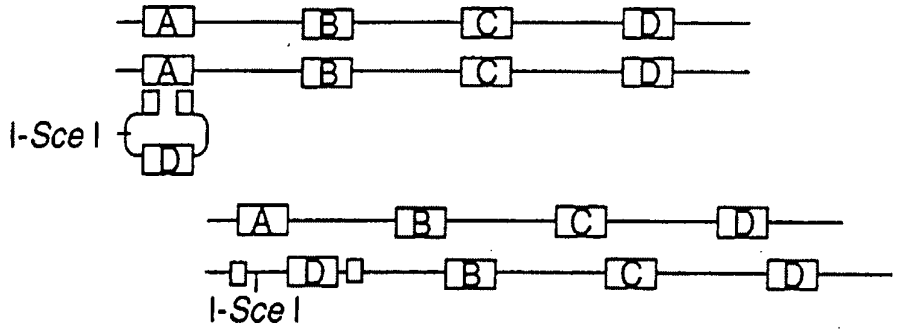


FIG. 30

DUPLICATION OF A LOCUS

1. INSERTION OF I-*Sce* I SITE BY CLASSICAL GENE REPLACEMENT



2. SPECIFIC CLEAVAGE BY I-*Sce* I ENZYME AND REPAIR OF THE BREAK BY HOMOLOGOUS SEQUENCES

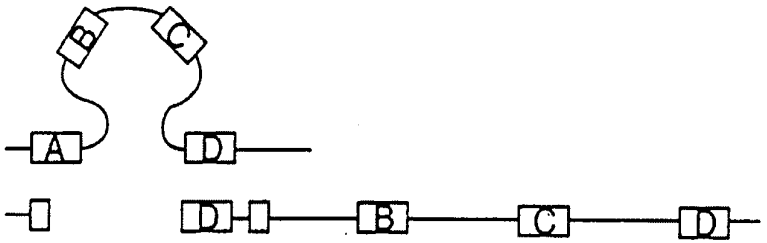
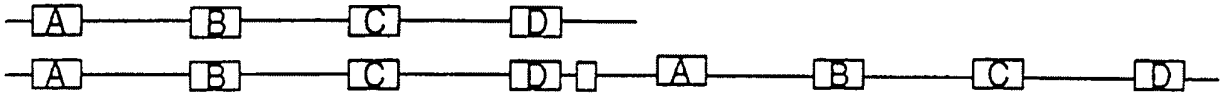


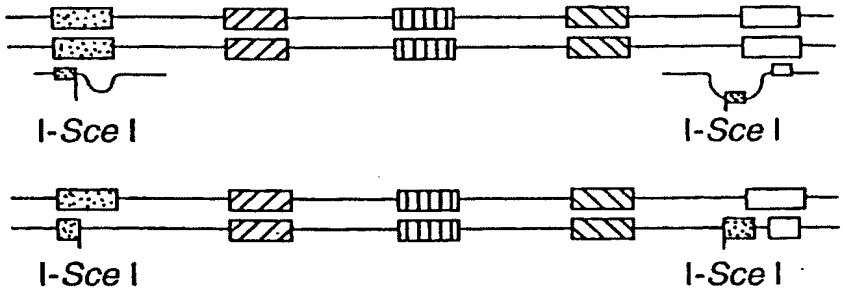
FIG. 31

3. DUPLICATION OF THE TOTALITY OF THE LOCUS

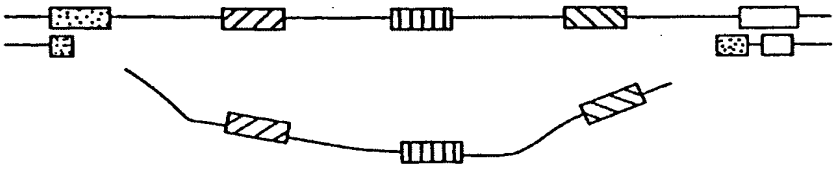


DELETON OF A LOCUS

1 INSERTION OF TWO I-*Sce* I SITES FLANKING THE LOCUS



2 EXPRESSION OF THE ENZYME AND CLEAVAGE



3 RECOMBINATION BETWEEN THE TWO ENDS

4 DELETION OF THE LOCUS

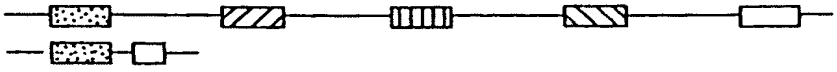


FIG. 32

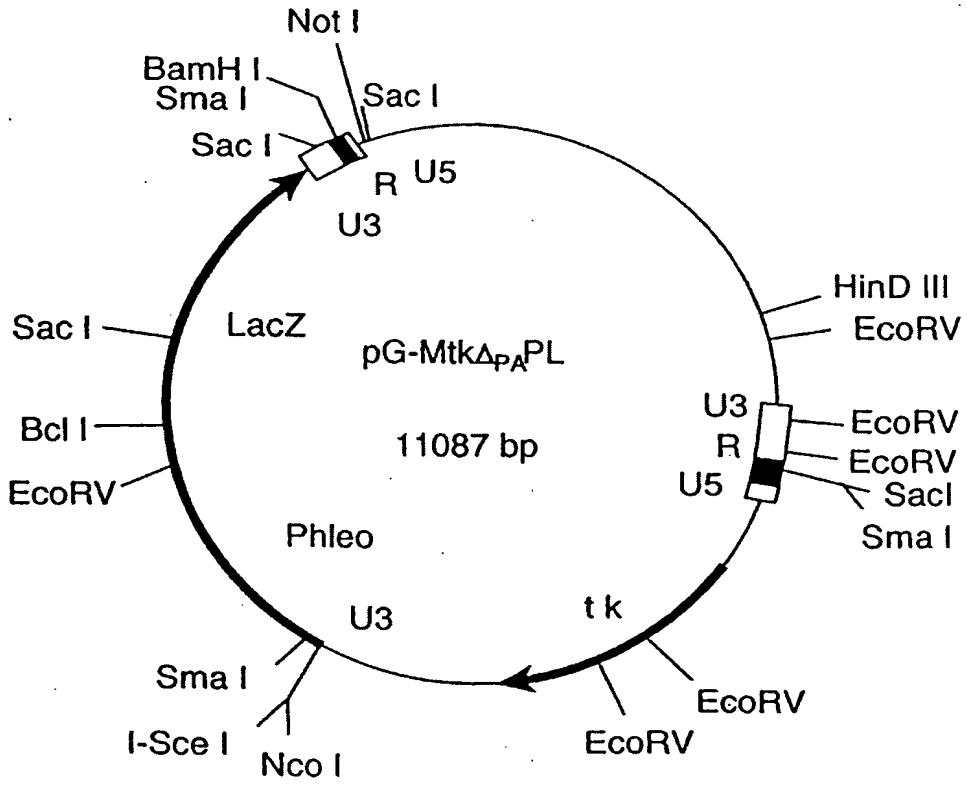


FIG. 33